Journal of Experimental Psychology

ARTHUR W. MELTON, Editor
Department of Psychology, University of Michigan
Ann Arbor, Michigan

DAVID A. GRANT, Associate Editor
University of Wisconsin

DELOS D. WICKENS, Associate Editor
Ohio State University

Consulting Editors

E. JAMES ARCHER, *University of Wisconsin*
JUDSON S. BROWN, *State University of Iowa*
CLETUS J. BURKE, *Indiana University*
PAUL M. FITTS, *Ohio State University*
FREDERICK C. FRICK, *Massachusetts Institute of Technology*
FRANK A. GELDARD, *University of Virginia*
JAMES J. GIBSON, *Cornell University*
CLARENCE H. GRAHAM, *Columbia University*
HAROLD W. HAKE, *University of Illinois*
LLOYD G. HUMPHREYS, *University of Illinois*
ARTHUR L. IRION, *Tulane University*
HOWARD H. KENDLER, *New York University*
DONALD B. LINDSLEY, *University of California, Los Angeles*
KENNETH MACCORQUODALE, *University of Minnesota*
QUINN MCNEMAR, *Stanford University*
L. STARLING REID, *University of Virginia*
KENNETH W. SPENCE, *State University of Iowa*

CONTENTS

Learning and Performance in a Tracking Task Under Two Levels of Achievement Information Feedback: A. F. SMODE 297
Effects of Discrimination Performance of Similarity of Previously Acquired Stimulus Names: K. J. NORCROSS 305
A Study of "Fittingness" of Signs to Words by Means of the Semantic Differential: G. A. MCMURRAY 310
Inhibition With Reinforcement (Conditioned Inhibition): D. C. KENDRICK 313
Extinction and Spontaneous Recovery of Conditioned Eyelid Responses as a Function of Amount of Acquisition and Extinction Training: W. F. PROKASY 319
Pursuit Rotor Performance as a Function of Delay of Information Feedback: E. J. ARCHER AND G. A. NAMIKAS 325
A Scale of Apparent Intensity of Electric Shock: S. S. STEVENS, A. S. CARTON, AND G. M. SHICKMAN 328
Reliability of Judgments of Figural Complexity: L. L. ELLIOTT 335
Generalization of an Instrumental Response With Variation in Two Attributes of the CS: S. H. WHITE 339
Facilitation of Mirror-Image Word Identification by Mirror-Image Perceptual Set: S. S. CULBERT 344
The Effect of a Stimulant and a Depressant Drug on a Measure of Reactive Inhibition: S. N. SINHA, C. M. FRANKS, AND P. L. BROADHURST 349
The Magnitude of Binocular Summation as a Function of the Method of Stimulus Presentation: G. COLLIER AND P. KUBZANSKY 355
Amount of Information Gained During Brief Exposures of Numerals and Colors: N. S. ANDERSON AND P. M. FITTS 362
Discrepancy From Adaptation Level as a Source of Affect: R. N. HABER 370
Anagram Solution Times: A Function of Letter Order and Word Frequency: M. S. MAYZNER AND M. E. TRESSELT 376
Absolute Judgments of Static Perspective Transformations: R. C. SIDOROKY 380

American Psychological Association

Vol. 56, No. 4 October 1958

ARTHUR C. HOFFMAN, Managing Editor
HELEN ORR, Promotion Manager
SADIE J. DOYLE, Editorial Assistant

The JOURNAL OF EXPERIMENTAL PSYCHOLOGY is published monthly, two volumes per year, by the American Psychological Association, Inc., at Prince and Lemon Sts., Lancaster, Pa. The subscription rate per volume is $8.00, or $16.00 annually. Single copies are $1.50. Subscriptions, orders, address changes, and business communications should be addressed to the American Psychological Association, Inc., 1333 Sixteenth St. N.W., Washington 6, D. C.

This Journal publishes original experimental investigations which are intended to contribute toward the development of psychology as an experimental science. Studies with normal human subjects are favored over studies involving abnormal or animal subjects, except when the latter are specifically oriented toward the extension of behavior theory. Experimental psychometric studies and studies in applied experimental psychology or engineering psychology may be accepted if they have broad implications for experimental and theoretical psychology.

Normally, articles exceeding 20 printed pages cannot be accepted. Within this limit, the piecemeal experiment-by-experiment reporting of psychological research is discouraged, and the reporting of the data and conclusions of substantial segments of research efforts is encouraged. Specifically, an integrated series of studies accomplished simultaneously (e.g., most doctoral dissertations) *must* be presented in a single article. Special provision is made for the publication of Supplementary Reports and Replication Reports in articles of not more than 1.5 printed pages (see Editor's Note, this Journal, 1957, 53, 1–2).

Address all articles submitted for publication to the Editor, Arthur W. Melton, Department of Psychology, University of Michigan, Ann Arbor, Michigan. **Manuscripts *must* adhere to the conventions concerning reference citations, preparation of tables and figures, manuscript format, etc., as described in the *Publication Manual of the American Psychological Association*.** When in doubt about practices of the Journal, authors should examine a recent issue.

Articles are published in the order of their receipt, except in rare circumstances. Authors are supplied with 50 free offprints without covers. All of the cost of an author's alterations in galley proof is charged to the author. Priority in publication is given to articles whose authors assume the full cost of publication. In 1958 the publication cost is $20 per printed page. Authors of priority publications receive no free offprints.

To save printing costs, supplementary material to articles in this Journal are deposited with the American Documentation Institute. Copies may be obtained by ordering the Document Number given in the footnote reference to ADI. Orders should be sent to ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D. C. Orders *must* be accompanied by checks or money orders payable to the Chief, Photoduplication Service, Library of Congress.

Entered as second-class matter, February 6, 1937, at the post office at Lancaster, Pa., under the Act of March 3, 1879. Acceptance for mailing at the special rate of postage provided for in paragraph (d-2), Section 34.40, P. L. and R. of 1948, authorized September 11, 1947.

Send address changes to 1333 Sixteenth St. N.W., Washington 6, D. C. Address changes must reach the Subscription Office by the 10th of the month to take effect the following month. Undelivered copies resulting from address changes will not be replaced; subscribers should notify the post office that they will guarantee second-class forwarding postage. Other claims for undelivered copies must be made within four months of publication.

© 1958 by the American Psychological Association, Inc.

Journal of
Experimental Psychology

VOL. 56, NO. 4 OCTOBER, 1958

LEARNING AND PERFORMANCE IN A TRACKING TASK UNDER TWO LEVELS OF ACHIEVEMENT INFORMATION FEEDBACK [1]

ALFRED F. SMODE [2]

Ohio State University

Information feedback or knowledge of performance can take many forms in a skill learning task. Miller (**21**) distinguishes between learning and action feedback, and Annett and Kay (**4**) distinguish between intrinsic feedback (such as proprioceptive stimulation resulting from a motor response) and augmented feedback (such as that provided by *E* or by a feedback circuit from a machine). The present experiment deals with the effects of achievement information feedback. This is the type of feedback which tells *S* how the results of his responses conform to some objective reference, and which is believed to be especially effective in increasing motivation.

Generally, it has been found that achievement information feedback (*a*) increases the level of performance in tasks that are highly overlearned (**5, 7, 13, 14, 20**), (*b*) increases the rate of improvement early in practice on a new task (**3, 6, 9, 10, 11, 17, 19, 23, 24, 25, 27, 29**), and (*c*) increases the frequency of reports that tasks are more interesting and less fatiguing (**2, 5, 9, 19, 23, 30**), as compared with conditions in which achievement information is withheld. These three features, considered together, have led to the inference that knowledge of achievement has a motivating effect on behavior. It is not at all clear, however, whether the more rapid improvement resulting from information feedback during the early practice trials on a task is indicative of more rapid learning, or only of a heightened level of performance attributable to greater motivation and effort.

Industrial research has shown that motivation to perform is related to job content, and to the methods and procedures employed in the task (**15, 16, 18, 22**). In the present investigation it was assumed that,

[1] This report is a condensation of a thesis submitted to the OSU Graduate School in partial fulfillment of the requirements for the Ph.D. degree. The writer is indebted to Paul M. Fitts under whose direction this investigation was conducted. The research was carried out in the OSU Laboratory of Aviation Psychology and was supported by the Air Force Personnel and Training Research Center under Contract No. AF 41(657)-70 with the OSU Research Foundation. Appreciation is also expressed to Earl A. Alluisi and Peter Gain for many valuable suggestions

[2] Now with Dunlap and Associates, Stamford, Connecticut.

under a higher feedback-of-information schedule, performance would be superior to that under a lower information-feedback schedule, due to motivational effects.

The theoretical position taken by the majority of writers is that information feedback has a large effect on task performance but may have little or no effect on learning, as indicated by retention, resistance to extinction, and similar measures (**12, 28**). However, it is difficult to elaborate an experimental design that is adequate in assessing performance and learning effects independently as a function of the achievement information available during training. The present study is intended as a contribution to this problem.

The evidence also is inconclusive concerning what happens when feedback information is withdrawn at some stage in learning. Some data indicate that performance under a high-level information schedule shows a decrement when the information level is decreased or withdrawn (**7, 9, 19**). Other data indicate that the superiority of groups trained under a high-level information schedule is maintained after the feedback information schedule is diminished or withdrawn (**20, 24, 25**).

Method

Tracking apparatus.—A one-dimensional electronic compensatory tracking apparatus was employed in this study. The tracking input was generated by a cam, driven by a constant-speed motor. Rotation of the cam actuated a rocker arm to change a potentiometer which, as part of an electrical bridge circuit, caused needle deflections from a null position on the tracking display. The input voltage variation was ±6 v.

The tracking display was a zero-center voltmeter with a range of ±6 v. Needle deflection was 13.3° per volt. The needle movement was slightly underdamped so that the needle oscillated several times if subjected to sudden large voltage changes. Positional control of the needle excursions in the tracking task was provided by a control knob.

Two scoring circuits, providing time-on-target and absolute integrated error scores, were incorporated into the apparatus. For time-on-target scoring, the error voltage was fed into two voltage comparator circuits which defined the scoring zone of the problem. A scoring zone bandwidth was set at 10° of arc, 5° on each side of the null position, corresponding to 6.25% of the input voltage (±.375 v.). Whenever the error, defined in terms of voltage differences between the input signal and *S*'s control signal, was within the specified scoring bandwidth, a "Sodeco" four-digit counter (Type TCeZ4E25) was activated, and cumulated an on-target score at the rate of 20 counts per second. This counter was mounted in a soundproof box. For absolute integrated error ($\int |e|$) scoring, the error voltage was fed into a conventional computer amplifier which translated error into absolute values which were then integrated and displayed on a meter at *E*'s station.

Two different cam-generated tracking inputs were used. The wave form of these inputs corresponded to a fundamental with a 6-cpm frequency and a second harmonic with a 12-cpm frequency. The amplitude ratio of these two components was 12:7.5 units. Tracking input No. 1 can thus be specified as:

$$\theta = (12 \sin\omega t + 7.5 \cos 2\omega t)\, A$$

where θ is angular degrees of needle deflection on the display, ω is .6288 radian/sec., t is time in seconds, and A is 3.07 (a constant). Tracking input No. 2, which was used by half the *S*s in the transfer phase of the study, was identical with input No. 1 with the exception that the second harmonic was phase-shifted a minus 90° from that in the first input. Problem input No. 2 can thus be specified as:

$$\theta = (12 \sin\omega t + 7.5 \sin 2\omega t)\, A$$

These inputs were chosen in the hope that they would provide two somewhat different tracking problems, comparable in difficulty level.

The tracking display was a meter, 3¼ in. in diameter. It was located approximately 24 in. in front of *S*, normal to his line of sight, and at eye level. A 3½-in. diameter control knob was located directly below the display.

Apparatus for providing achievement information feedback.—A decade counter[3] was used for presenting additional visual knowledge of

[3] The idea of using a decade counter plus auditory clicks originated from work supported by a grant from the OSU Graduate School, in collaboration with Paul M. Fitts and J. Alfred Leonard.

achievement information. It was butted at a 45° angle to the right of the tracking apparatus. It consisted of three columns of red 1¼-in. panel lights mounted on a 10½ × 17-in. black aluminum panel. Each column of lights was numbered upwards consecutively from 0 to 9. The lights in the right column were valued at one unit while the middle and left column lights were valued at 10 and 100 units, respectively. The lights were sequenced by stepping switches and relays and made possible the visual display of 999 score units, each unit representing .5 sec. time-on-target. Groups not using the decade counter were not made aware of its function in the task and regarded it simply as some necessary component in the "complex" apparatus; it was turned off in this case.

Clicks heard in a set of earphones provided additional auditory knowledge of achievement information. Clicks occurred at the rate of 2 per sec. as long as *S* was on target.

Procedure.—The same initial instructions were read to each *S*. He was told that when the problem began he should move the control knob so as to try to keep the needle of the tracking display on the zero mark at all times. The training task was identical for all *S*s except that the knowledge of achievement feedback was varied. Additional specific instructions were read to each *S*, depending upon the information feedback schedule assigned (H or L) and upon transfer to the same or a new target course. Each *S* served for a 1-hr. session of 14 trials, of which 11 were training trials and three were transfer trials. A 1-min. rest period was given between all trials except for a 3-min. rest between Trial 11, the last training trial, and Trial 12, the first transfer trial. The 3-min. rest period allowed *E* time to change problem input cams manually when required. Each trial lasted for 90 sec., of which the first 5 sec. were not scored. This 5-sec. warm-up period gave *S* an opportunity to center the needle in the on-target zone before the scoring started.

The high-level knowledge of achievement groups (H-Groups) received clicks in the earphones at the rate of one for each .5 sec. of cumulated time-on-target, and for each .5 sec. *S* remained on the target the next higher red light was also lit on the decade counter. The low-level knowledge of achievement groups (L-Groups) received broad band noise in the earphones of sufficient intensity to mask the clicking of the scoring relays which would otherwise have provided auditory performance cues. At the end of each trial they were told their time-on-target score for that trial to the nearest second. It can be seen that the H-Groups received a greater amount of information per trial than the L-Groups. The information was presented immediately to the H-Groups, as accrued, in small incremental steps, whereas the L-Group information was presented only after each trial was completed, through a different mode of presentation. It should be noted that the two schedules of information feedback are defined in terms of the aggregate (confounded) effect of the basic high and low information feedback conditions which differed with respect to (*a*) the amount of achievement information presented, (*b*) the sensory mode of information presentation, and (*c*) the temporal character of presentation. It was not a purpose of the study to assign quantitative values to the parameters in each condition. It should be noted that both groups could observe their own performance error continually on the tracking display; however, there was no objective standard available on the display by means of which they could evaluate the goodness of this performance or their cumulative achievement. For all groups, time-on-target scores were recorded from the Sodeco counter and integrated error scores from the scoring circuit.

In the transfer phase of the study eight conditions were differentiated—one-half of the *S*s continued on the information schedule imposed at the outset, and one-half changed over to the other basic information schedule; further, one-half the *S*s continued to track target input No. 1 (used during the training phase), and one-half transferred to target input No. 2. The eight subgroups are shown in Table 1.

Subjects.—The *S*s were 160 male volunteer undergraduate students from a beginning course in psychology at the Ohio State University. As *S*s reported, they were assigned randomly to

TABLE 1

THE VARIOUS KNOWLEDGE OF ACHIEVEMENT (K OF P) AND TARGET INPUT (TASK 1 OR TASK 2) CONDITIONS EMPLOYED DURING TRAINING AND TRANSFER CONDITIONS

Group	Training (Trials 1–11)		Transfer (Trials 12–14)	
	K of P	Task	K of P	Task
1	High	1	High	1
2	High	1	High	2
3	High	1	Low	1
4	High	1	Low	2
5	Low	1	High	1
6	Low	1	High	2
7	Low	1	Low	1
8	Low	1	Low	2

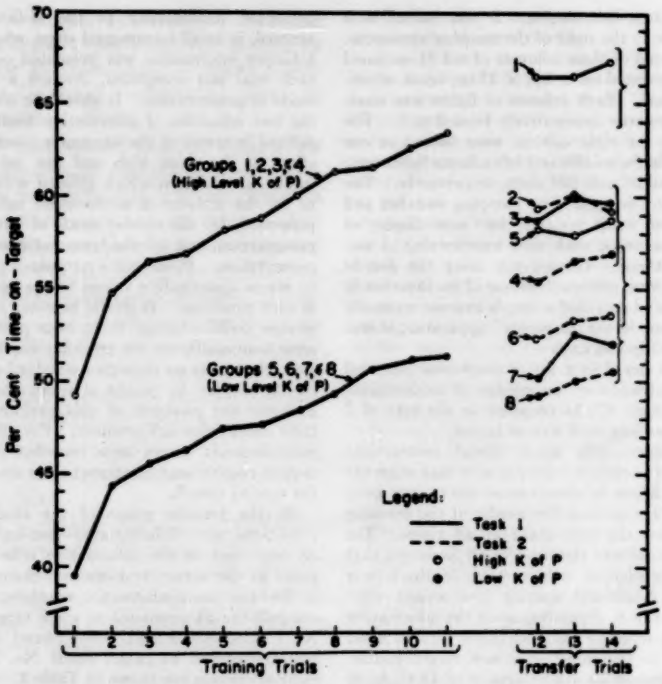

FIG. 1. Percentage time-on-target scores for the high-level and low-level knowledge of achievement (K of P) groups over training and transfer trials (20 *Ss* per group). Brackets indicate groups that did not differ significantly on the three transfer trials (Duncan's Multiple Range Test).

one of the eight experimental subgroups in such a way as to keep the numbers in the groups as nearly balanced as possible, until each subgroup contained 20 *Ss*.

Protocols.—In addition to the time-on-target and integrated error scores, a protocol of subjective reactions to the tracking task was obtained from each *S*. After completing the experimental session, each *S* was asked: "What is your reaction to this kind of task?" "How does it strike you with regard to interest, monotony, or fatigue?" Tabulation of the answers to these questions provides a basis for a comparison of subjective reactions to the task under the two conditions of information feedback.

RESULTS

Time-on-target scores.—Percentage time-on-target scores are shown in Fig. 1. For the time-on-target data, Bartlett's test did not preclude the assumption of homogeneity of variance. A t test between the performance of the H- and L-Groups ($N = 80$ per group) over the first 11 training trials was highly significant ($P < .01$), the H-Group being superior. Both groups showed consistent improvement over the training trials; however the H-Groups were superior on the first 90-sec. trial and maintained this superiority in performance throughout all 11 training trials. It is apparent that the augmented achievement information feedback facilitated performance.

A $2 \times 2 \times 2$ factorial analysis of variance was carried out for the

data of Trials 12, 13, and 14 combined, in order to assess the experimental effects in terms of performance on the transfer trials. These results are summarized in Table 2. Significant variation ($P < .01$) obtained among the means of the three major variables of original information feedback (OI), transfer information feedback (TI), and target course (C). In addition, the interaction of transfer information feedback and target course was also significant ($P < .05$).

The TI × C interaction is interpreted as indicating that the joint effect of high-level information feedback plus tracking of the same target input, as contrasted with the joint effect of a low-level feedback plus tracking of a different target input, was greater over transfer trials than could be predicted from the main effects of transfer information feedback (TI) and target course (C) considered separately.

A Duncan Multiple Range Test (8) was carried out to determine nonsignificant differences in performance among the eight experimental groups over the combined transfer trials. The critical comparisons for the present study are those between groups trained under different information feedback conditions and tested under identical feedback conditions on the transfer trials. Four such comparisons of transfer performance are possible (Groups 1 vs. 5, 2 vs. 6, 3 vs. 7, and 4 vs. 8). All four comparisons revealed significant differences favoring the group trained under the condition of high achievement information feedback.

Also of special interest is the comparison of the transfer performance of high-to-low vs. low-to-high information feedback groups (Groups 3 vs. 5 and 4 vs. 6). In both comparisons the high-to-low group was superior, but neither difference was statistically significant. This finding suggests, for the present task and training conditions, that the facilitating effects of high information feedback on learning during early trials is approximately balanced by the facilitating effects of high feedback on performance during the transfer trials, i.e., that learning and performance effects were approximately equal in this task.

TABLE 2

ANALYSIS OF VARIANCE FOR PERCENTAGE TIME-ON-TARGET FOR ALL GROUPS OVER COMBINED TRANSFER TRIALS 12–14

Source	*df*	*F*	*P*
Between groups	7	15.6599	.01
Original information feedback (OI)	1	56.6534	.01
Transfer information feedback (TI)	1	28.3053	.01
Target course (C)	1	19.7881	.01
OI × TI	1	.0294	
OI × C	1	.6423	
TI × C	1	4.1848	.05
OI × TI × C	1	.0162	
Within groups	152	(106.4709)[a]	
Total	159		

[a] Mean square for error.

Integrated error scores.—For the absolute integrated error data, the results of Bartlett's test did not warrant the assumption of homogeneity of variance. Accordingly, a nonparametric analysis (1) was carried out for the pooled data of the three transfer trials in order to assess the experimental effects among the three variables. The results are very substantially in agreement with those for the time-on-target scores. The ranks of the eight groups in terms of integrated error scores were 1, 2, 3, 4, 5, 6, 8, 7, which agrees very closely with their rankings in terms of the time-on-target scores. A significant difference ($P < .01$) was shown for the effects of high vs. low training information feedback, and also ($P < .01$) for the transfer information feedback variable. The effect of the target course (C) variable, however, was not significant and no significant interaction effects were found. Apparently the nonparametric analysis of integrated error did not provide as

sensitive a test of performance differences as did the parametric analysis of time-on-target scores.

Protocols.—An analysis of *S*s' subjective reactions to the tracking task indicated that interest level accrued as a function of increased information feedback. Most *S*s who served under H conditions reported that the decade counter and the auditory clicks seemed to produce a high level of interest in the task. Of the 40 *S*s (Subgroups 7 and 8) who performed under a low information feedback throughout, 37.5% stated that the task was "interesting," 25% said the task was "average" in interest level, and 37.5% stated that the task was "not interesting." It is important to note that of the small number of *S*s in the L-Groups who judged the task interesting, two-thirds added some kind of qualification to the effect that the interest level fell off and the task became boring after several trials. In contrast to this, of the two groups ($N = 40$) that performed under H conditions throughout (Subgroups 1 and 2), 87.5% stated without qualification that the task was an interesting one, 7.5% said the task was average in interest level, and only 5% stated that the task was not interesting. A chi-square analysis of the preceding data indicated a difference between the H and L groups that was significant at the $P < .01$ level. Four experimental groups ($N = 80$) performed under both schedules of performance information and were therefore able to compare the two conditions. Of this group, 80% preferred high information feedback, 15% had no preference, and only 5% preferred low information feedback.

An analysis of the contents of the individual protocols suggests three bases for the motivation effects: (*a*) *feelings of tension*—under the high-level information condition, some *S*s reported a general increase in attentiveness to the task together with feelings of increased tension and pressure to perform well; (*b*) *self-competition*—*S*s cited the competitive value of the decade counter and its high incentive value, and gave statements framed in terms of trying for a high-score goal; (*c*) *game-playing attitude*—some *S*s reported that the decade counter made the task easier and more fun, engendering a better attitude toward the task, so that they viewed the task as a game such as playing a pinball machine.

Discussion

Three theoretical issues are raised by these results. (*a*) Do the behavioral effects resulting from the higher information feedback condition arise directly from the utilization of additional information regarding task performance, or are they mediated indirectly through an increase in motivation? (*b*) What is the basis of the hypothesized increase in motivation? (*c*) Is the learning effect noted on transfer trials due to the fact that tracking skills are increased, or to the fact that *S*s have developed better work habits? These issues are progressively more difficult to answer on the basis of the present data.

Regarding the first issue, it has been shown by numerous investigators that the type of display used in a tracking task (pursuit vs. compensatory, quickened vs. unquickened, etc.) affects the level of performance attained by *S*s. However, in the present task the primary tracking display was the same for all groups. Furthermore, had *S*s attempted to track by means of the additional information feedback provided by the clicks and the decade counter, they would have found it very difficult or impossible since this information was not sufficiently specific as to the direction and magnitude of error to permit *S* to use it as a cue to the making of specific responses. In fact, there is even reason to speculate that the decade counter may have been in one sense a distraction. Therefore, we can safely conclude that the effect of the higher information feedback was mediated by an increase in motivation.

Regarding the second issue, of why the increased motivation, it was stated earlier that it was not the purpose of this experiment to investigate this problem in terms of the several characteristics of the heightened feedback information schedule. However, it is hypothesized that a schedule providing for the frequent feedback of achievement information, categorized as good or bad against an external standard, has generally a motivating as well as a reinforcing effect. Such an effect has been found in two other (unpublished) studies conducted in the Laboratory of Aviation Psychology, which employed a very different (perceptual discrimination) learning task.

The third issue cannot be adequately settled on the basis of the present data. On the one hand, it is possible that *S*s learned more about the specific tracking

task while performing at a high level during training trials; on the other hand, they may have developed better work habits, such as attending more closely to the display, making quicker and more frequent corrections, etc., and these more efficient work habits, or the attitudes associated with them, may have carried over to the transfer trials. The present writer prefers to identify the effects as learning; however, the alternative carry-over-of-motivation hypothesis cannot be ruled out. It is quite possible, of course, that both types of effects were present—for example, that a higher rate of responding characterized both training and transfer trials.

The type of feedback information provided here appears to have some of the characteristics of the reinforcement schedules found most effective by Skinner (**26**) in producing high rates of responding. It may be fruitful to investigate the relation of augmented information feedback schedules to rate of responding.

Summary

The present experiment was designed to provide an independent assessment of performance effects and learning effects in a compensatory tracking task as a function of the method used in providing cumulative information as to achievement level. With the exception of the procedural variable of information feedback (knowledge of achievement) the task was uniform for all *Ss*. A transfer of training design was employed, differentiating two basic conditions (high and low information feedback schedules) of the training phase into eight experimental subgroups in the transfer phase. On the transfer trials, one-half of the *Ss* continued to receive the same type of information imposed at the outset and one-half changed to the other schedule; one-half continued with the original target course while one-half transferred to a new target course equated for difficulty level. The high and low information feedback conditions differed in terms of the aggregate effect of the following three parameters: (*a*) the amount of information presented, (*b*) the sensory mode of presentation, and (*c*) the temporal characteristics of presentation.

For both the time-on-target and the absolute integrated error scores, all groups showed consistent improvement over training trials; however, the high information feedback groups were superior at the end of the first 90-sec. trial, and maintained this superiority throughout the training period. Thus, the high-level information feedback clearly facilitated performance.

Comparisons of groups trained under different information feedback conditions and tested under identical conditions on transfer trials revealed significant differences which in all cases favored training under high-level information feedback. These are identified as learning effects, although a carry-over-of-motivation hypothesis (persistence of a favorable attitude) cannot be ruled out.

For the present task and training conditions, the facilitating effects of high information feedback during early training trials (as measured by the differences between control and experimental groups under low information feedback) were approximately balanced by the facilitating effects of high information feedback during transfer trials (as measured by transfer performance under high information feedback).

Protocols of subjective reactions to the tracking tasks indicated that interest level accrued as a function of increased information feedback. Possible bases for this incentive differential are discussed. It is concluded that the effect of the higher information feedback was mediated by an increase in motivation.

It is suggested that the manipulation of extrinsic information feedback may prove to be a useful technique for controlling human motivational levels in a variety of learning tasks.

References

1. Alluisi, E. A. Computational formulae for a distribution-free test of analysis-of-variance hypotheses. *USAF WADC Tech. Rep., No. 56-339*, 1956.
2. Ammons, R. B. Effects of knowledge of performance: A survey and tentative theoretical formulation. *J. gen. Psychol.*, 1956, **54**, 279–299.
3. Angell, G. W. The effect of immediate knowledge of quiz results on final examination scores in freshman chemistry. *J. educ. Res.*, 1949, **42**, 391–394.
4. Annett, J., & Kay, H. Knowledge of results and skilled performance. *Occup. Psychol.*, 1957, **31**, 69–79.
5. Arps, G. F. Work with knowledge of results versus work without knowledge of results. *Psychol. Monogr.*, 1920, **28**, No. 3 (Whole No. 125).
6. Bilodeau, E. A. Some effects of various degrees of supplemental information given at two levels of practice upon the acquisition of a complex motor skill.

USAF Hum. Resour. Res. Cent. Res. Bull. No. 52-15, 1952.

7. Book, W. F., & Norvell, L. The will to learn: An experimental study of incentives in learning. *Ped. Sem.*, 1922, **29**, 305–362.
8. Duncan, D. B. Multiple range and multiple *F* tests. *Biometrics*, 1955, **11**, 1–42
9. Elwell, J. L., & Grindley, G. C. The effect of knowledge of results on learning and performance. I. A co-ordinated movement of the two hands. *Brit. J. Psychol.*, 1938, **29**, 39–53.
10. Gilliland, A. R. The effect of practice with and without knowledge of results in grading handwriting. *J. educ. Psychol.*, 1925, **16**, 532–536.
11. Hamilton, H. C. The effect of incentives on accuracy of discrimination measured on the Galton bar. *Arch. of Psychol.*, 1929, **16** (103), 1–73.
12. Hull, C. L. Behavior postulates and corollaries—1949. *Psychol Rev.*, 1950, **57**, 173–180.
13. Johanson, A. M. The influence of incentive and punishment upon reaction-time. *Arch. of Psychol.*, 1922, **8** (54).
14. Judd, C. H. Practice without knowledge of results. *Psychol. Res. Monogr. Suppl.*, 1905, **7**, 185–199.
15. Kahn, R., & Morse, N. The relationship of productivity to morale. *J. soc. Issues*, 1951, **7** (3), 8–17.
16. Katz, D., & Kahn, R. Human organization and worker motivation. *Industr. Productivity*, 1952. Pp. 146–171.
17. Lindahl, L. C. Movement analysis as an industrial training method. *J. appl. Psychol.*, 1945, **29**, 420–436.
18. Mace, C. A. Satisfactions in work. *Occup. Psychol.*, 1948, **22**, 5–19.
19. MacPherson, S. J., Dees, V., & Grindley, G. C. The effect of knowledge of results on learning and performance: II. Some characteristics of very simple skills. *Quart. J. exp. Psychol.*, 1948, **1**, 68–78.
20. Manzer, C. W. The effect of knowledge of output on muscular work. *J. exp. Psychol.*, 1935, **18**, 80–90.
21. Miller, R. B. Handbook on training and training equipment design. *USAF WADC Tech. Rep. No. 53-136*, 1953.
22. Morse, N. C. Satisfactions in the white-collar job. Ann Arbor: Surv. Res. Cent., Inst. Soc. Res., Univer. of Michigan, 1953.
23. Pressey, S. L. Development and appraisal of devices providing immediate automatic scoring of objective tests and concomitant self-instruction. *J. Psychol.*, 1950, **29**, 417–447.
24. Reynolds, B., & Adams, J. Motor performance as a function of click reinforcement. *J. exp. Psychol.*, 1953, **45**, 315–320.
25. Ross, C. C. An experiment in motivation. *J. educ. Psychol.*, 1927, **18**, 337–346.
26. Skinner, B. F. The experimental analysis of behavior. *Amer. Scientist*, 1957, **45**, 343–371.
27. Thorndike, E. L. The law of effect. *Amer. J. Psychol.*, 1927, **39**, 212–222.
28. Tolman, E. C. *Purposive behavior in animals and man*. New York: Century, 1932.
29. Trowbridge, M. H., & Cason, H. An experimental study of Thorndike's theory of learning. *J. gen. Psychol.*, 1932, **7**, 245–260.
30. Wright, W. R. Some effects of incentives on work and fatigue. *Psychol. Rev.*, 1906, **13**, 23–34.

(Received January 10, 1958)

Journal of Experimental Psychology
Vol. 56, No. 4, 1958

EFFECTS OF DISCRIMINATION PERFORMANCE OF SIMILARITY OF PREVIOUSLY ACQUIRED STIMULUS NAMES [1]

KATHRYN J. NORCROSS [2]

Iowa Child Welfare Research Station

If *S* learns discriminal verbal responses to a given set of similar stimuli in one task and is subsequently required to learn motor responses to these same stimuli presented in a transfer task, performance on that transfer task has been shown to be facilitated (6). It has been suggested that when verbal responses are learned to a set of environmental stimuli, the response-produced stimulation accompanying verbalization will contribute to each total stimulus complex. In conjunction with other principles in learning theory, facilitation in a transfer task may be predicted as an increasing function of the distinctiveness of the response-produced stimuli. On the assumption that the similarity of the responses is correlated with the similarity of the stimuli they produce, it would be predicted that *S*s who learn similar names for the experimental stimuli will not do as well on the subsequent criterion task as *S*s who learn distinctive names, since the effective stimulus complexes in the criterion task will be more distinctive the more dissimilar the responses learned in pretraining.

Pretraining experience involving one of two sets of stimulus names, either three nonsense syllables judged similar or three nonsense syllables judged dissimilar, was given by Gerjuoy (1). No advantage was indicated in the transfer task for the use of dissimilar names as compared to the use of similar names. The present experiments proposed to investigate further the effects of prior attachment of names of differing degrees of similarity to the stimuli to be discriminated in a transfer task. A within-Ss design was used in an attempt to control extraneous variables which are known to be factors in transfer, as well as to increase the precision of the experiments.

Experiment I

Method

Design.—The experimental design is schematized in Table 1. Experimental stimuli were photographs of pen and ink line drawings of Indian children's faces—a pair of girls' faces and a pair of boys' faces. In pretraining, *S*s learned highly similar nonsense names for one pair and dissimilar nonsense names for the other pair. Thirty kindergarten children [3] were randomly as-

TABLE 1

Experimental Design

Pretraining		Transfer Task
Subgroup 1	Subgroup 2	
Boy 1—zim	Boy 1—wug	Boy 1—Button 1
Boy 2—zam	Boy 2—kos	Boy 2—Button 2
Girl 1—wug	Girl 1—zim	Girl 1—Button 3
Girl 2—kos	Girl 2—zam	Girl 2—Button 4

[1] This research was the basis for a doctoral dissertation under the guidance and direction of Charles C. Spiker, who also contributed immeasurably to the paper in its present form.

[2] Now at Trenton State Teachers College, Trenton, New Jersey.

[3] Appreciation is expressed to Hale C. Reid, Director of Curriculum and Instruction, for permission to obtain subjects in the Cedar Rapids Public School. Thanks are also due to John Manville, Principal, and to the staff of Hayes School for administrative assistance and cooperation in the collection of data for this study.

signed to one of two subgroups used for purposes of counterbalancing. The transfer task, identical for all Ss, required *S* to push a different one of four buttons to each of the separately presented pictures. It was expected that the learning performance, in terms of correct responses, would be superior for the pair of stimuli with dissimilar names as compared to the pair with similar names.

Apparatus.—The apparatus used in transfer was comprised of two main units. The stimulus unit, an engineering model of the Hunter Cardmaster, has been previously used and described by Norcross and Spiker (5). Briefly, this visual stimulus presentation device consists of a gray metal box 14 × 14 × 14 in., with a 3 × 6-in. aperture in the front face. Plastic cards on which the stimulus pictures were mounted were presented before this aperture.

The response unit was below the stimulus aperture and in front of *S*. As seen by *S*, this consisted of a gray wooden box with the top panel sloping downward from the aperture. On this panel, four push-button switches were placed at what would be the corners of a 3-in. square. The two buttons appropriate for a given pair of pictures were diagonally located at the points of the square and were marked with colored tape—those for the Indian girls with yellow tape and those for the Indian boys with red tape. A small informational lamp was located immediately beside each button.

Procedure.—The name learning was administered on Day 1. The *E* verbalized the names and indicated which name went with each picture, for the girls' and for the boys' pictures. Some practice was then given, without the pictures, in serial recitation of the two names appropriate for each pair. At *E*'s discretion, the learning trials were begun in which a single picture was held up by *E* and *S* was asked for the name. When a name appropriate for a stimulus member of the other pair was given, *E* offered the names for the pair of which this picture was a member, and again asked for a response. If *S* did not give any name, *E* asked for a choice of the responses appropriate to the pair of which the picture was a member.

Following a correct response, the next picture was presented. If an incorrect name was given, *S* was first corrected and then shown the picture for which the name he had given was correct. The order was such that presentation of a boy's picture alternated regularly with presentation of a girl's picture. A trial refers to four successive presentations in which all pictures were shown. Pretraining on Day 1 was continued until a criterion was achieved of one trial in which *S* had correctly volunteered the names for all pictures. Pretraining was discontinued for any child who did not meet the criterion in 18 trials and his record was not considered.

On Day 2, one to three days later, the pictures were presented by means of the Cardmaster until the criterion of one errorless naming trial was again reached. Transfer learning was immediately begun. The *S* was asked to give the name and then push a button as each picture was presented. If a boy's picture was shown and either one of the boys' names was given, it was accepted. This was also true for the girls' pictures. If *S* gave a girl's name for a boy's picture, or vice versa, *E* reminded him of the appropriate pair of names.

Only one button response followed a picture presentation. After any pushing response, the informational light came on beside the button correct for that picture. A bell, located inside the response unit, was sounded if the correct button had been pushed. The *E* determined when a new picture was presented and which response would be appropriately rewarded or corrected on that presentation.

A block of four stimulus presentations in which all pictures were shown made up a trial. The order of presenting the stimuli was random with the restrictions that no one picture appear twice in succession and that the number of times three girls' pictures appeared in sequence equal the instances of a sequence of three boys' pictures. Fifteen transfer trials were administered, while all naming and pushing responses were recorded.

Results

Verbal learning.—An average of 7.37 trials, including criterial trials, was required to achieve criterion on Day 1. In those trials, the mean number of correct dissimilar-name responses was 8.67 and the mean number of correct similar-name responses was 7.40. The value of t obtained from a test of this difference was 1.09 ($P > .20$). Data of three Ss who did not meet criterion were not considered. An average of 3.17 relearning trials was given on Day 2. The mean difference between similar and dissimilar correct responses was .53 ($t = 1.70$, $P > .05$).

Motor Learning.—Preliminary analysis revealed that the subgroups used

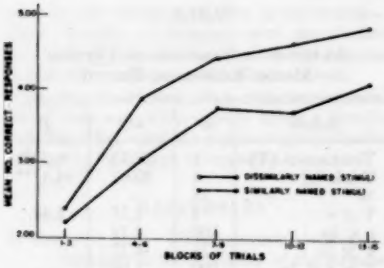

Fig. 1. Mean number of correct motor responses in Exp. I.

for counterbalancing did not differ in performance on the motor task and that they did not interact with experimental conditions. Therefore, the subgroup classification was disregarded in further analyses. For all *Ss* combined, a mean of 20.20 correct motor responses followed presentation of those stimuli which possessed dissimilar names while a mean of 16.90 correct responses followed presentation of the stimuli with similar names. Figure 1 graphically shows the means of the number of correct responses in blocks of three trials for the two similarity conditions separately. A Treatment × Trials × *Ss* analysis (4) of correct responses is summarized in Table 2.

TABLE 2

ANALYSIS OF VARIANCE OF CORRECT MOTOR RESPONSES, EXP. I

Source	*df*	*MS*	*F*
Treatments (T)	1	32.67	7.67*
Trials (t)	4	44.12	23.85**
Ss	29		
T × t	4	1.26	<1.00
t × *Ss*	116	1.85	
T × *Ss*	29	4.26	
T × t × *Ss*	116	1.50	
Total	299		

* $P < .01$.
** $P < .001$.

Discussion

The results obtained confirmed the prediction that more correct responses would be made in the learning of discriminal motor responses to a pair of stimuli for which dissimilar names had been previously learned, as compared to a pair for which similar names had been learned. However, an examination of the verbal responses occurring in the motor task revealed that 103 incorrect dissimilar verbal responses were given while 341 incorrect similar responses were made. The mean difference of 8.27 is significant at the .001 level. The explanation could be offered that a greater number of motor errors to the similarly named pair of stimuli occurred because the similar verbal cues were more inconsistently associated with their respective experimental stimuli than was true for the dissimilar verbal cues, and were therefore more irregularly reinforced with regard to the button responses. A second experiment was conducted which assured for each picture a consistent verbal response immediately prior to the motor response.

EXPERIMENT II

Experiment II differed from Exp. I only in that *S* was corrected for misnaming in the transfer task, and was required to verbalize the correct name prior to the button-pressing response. Twenty-six *Ss* were randomly assigned to one of the two subgroups.

Results

Verbal learning.—The mean number of trials on Day 1 was 7.54. An average of 3.31 relearning trials was required on Day 2. On neither day was there a significant difference between the number of correct similar and dissimilar responses ($P > .20$). Data of three *Ss* who did not meet criterion were not considered.

Motor learning.—A mean of 22.73

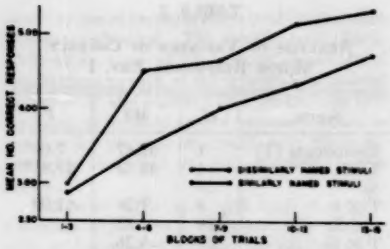

FIG. 2. Mean number of correct motor responses in Exp. II.

correct responses occurred to the pair with dissimilar names while a mean of 19.62 correct responses occurred to the pair with similar names. Table 3 summarizes a Treatment × Trials × *Ss* analysis of variance which again yielded both significant Treatment and Trials effects. Figure 2 depicts the results graphically.

TABLE 3

ANALYSIS OF VARIANCE OF CORRECT MOTOR RESPONSES, EXP. II

Source	*df*	*MS*	*F*
Treatments (T)	1	28.45	6.51*
Trials (t)	4	80.67	34.47**
Ss	25		
T × t	4	1.55	1.44
t × *Ss*	100	2.34	
T × *Ss*	25	4.37	
T × t × *Ss*	100	1.08	
Total	259		

* $P < .025$.
** $P < .001$.

Discussion

Experiment II showed a significant difference in transfer performance for the two types of stimuli with names of differing degrees of similarity, a confirmation of Exp. I. These experiments were designed so that each *S* served in both experimental conditions. Thus, there should be no differential effects of warm-up and learning-to-learn, both known to be nonspecific factors in transfer of training (2, 7). Each *S* had the same amount of experience in learning to respond discriminally to the two pairs in pretraining. Therefore, any effects established by this experience, e.g., "observing responses" (3), should have transferred to the second task equally for both pairs of stimuli. It would not seem that any of these factors can be used to account for differences in performance for the two similarity conditions. In addition, Exp. II controlled verbalization so that *S* consistently gave the correct name immediately prior to the motor response. Thus it appears that it is the difference in similarity of the pairs of verbal responses which has differentially influenced performance on the criterion task.

It should be noted that the design of these experiments does not permit inference as to whether the transfer was negative, positive, or both. In order to determine the direction of transfer in such an experiment a control condition is needed which involves criterion task stimuli for which the *Ss* have not previously learned names.

It has previously been shown that performance when learning motor responses to a set of stimuli will be affected if verbal responses have been learned to these same stimuli. This study suggests that, under certain conditions, the transfer effects depend at least in part upon the degree of generalization among the response-produced stimuli.

SUMMARY

This study measured the effects on discrimination performance in kindergarten *Ss* of the degree of similarity of previously acquired stimulus names. In Exp. I, *Ss* learned similar nonsense names to two stimuli and dissimilar nonsense names to two other stimuli. These same stimuli were then shown in a transfer task in which *S* was required to give a name and push a different one of four buttons in response to each stimulus singly presented. More incorrect names and fewer correct motor responses occurred to the similarly named pair as compared to the dissimilarly named pair.

Experiment II differed from Exp. I only in that incorrect verbalizations were corrected

prior to the occurrence of a motor response. Again, transfer performance with the dissimilarly named pictures was significantly superior to that for the pictures with similar names.

It was concluded that response-produced verbal cues may differentially affect a transfer task depending upon the degree of generalization among the verbal cue components.

References

1. Gerjuoy, I. R. Discrimination learning as a function of the similarity of the stimulus names. Unpublished doctor's dissertation, State Univer. of Iowa, 1953.
2. Hamilton, C. E. The relationship between length of interval separating two learning tasks and performance on the second task. *J. exp. Psychol.*, 1950, **40**, 613–621.
3. Kurtz, K. H. Discrimination of complex stimuli: The relationships of training and test stimuli in transfer of discrimination. *J. exp. Psychol.*, 1955, **50**, 283–292.
4. Lindquist, E. F. *Design and analysis of experiments in psychology and education.* New York: Houghton Mifflin Co., 1953.
5. Norcross, K. J., & Spiker, C. C. The effects of mediated associations on transfer in paired-associate learning. *J. exp. Psychol.*, 1958, **55**, 129–134.
6. Spiker, C. C. Experiments with children on the hypotheses of acquired distinctiveness and equivalence of cues. *Child Develpm.*, 1956, **27**, 253–263.
7. Thune, L. E. The effect of different types of preliminary activities on subsequent learning of paired-associate material. *J. exp. Psychol.*, 1950, **40**, 423–438.

(Received September 6, 1957)

Journal of Experimental Psychology
Vol. 56, No. 4, 1958

A STUDY OF "FITTINGNESS" OF SIGNS TO WORDS BY MEANS OF THE SEMANTIC DIFFERENTIAL

GORDON A. McMURRAY

University of Saskatchewan

The term "fittingness" to describe the relationship between a word and a pictorial sign was used by Bartlett (1, Ch. 6) in reporting the results of his Picture Writing experiment. His *Ss* were required to associate certain word-sign pairs and to reproduce the correct sign whenever the corresponding word appeared in a short story dictated 15 min. after the learning period. Some word-sign combinations were accepted by his *Ss* as being peculiarly fitting, and the recognition of this fittingness evoked a pleasant feeling. The conditions which gave rise to this "fittingness" experience were not studied, but Bartlett considered that it probably involved complex cognitive processes for which a verbal formulation might or might not be possible.

Hall and Oldfield (3), using six words each matched with two signs, found significant agreement among *Ss* in the choice of the "best-fitting" signs. The reasons given for the choices were so varied that they appeared as rationalizations for a more basic process in which the general formal and structural qualities of the signs influenced the selections made. More recently, Hall (2) used 50 words each matched with five possible symbols. The agreement of choice for many words was again high with the basis of fittingness varying from signs having a direct conventional significance to those possessing some undefined formal quality which fitted the word.

Similar results have been obtained in experiments where different tasks were used. Usnadze (7) presented 42 unfamiliar syllables or sound complexes from which an appropriate name was to be selected for each of six drawings. Agreement of choice was again beyond that which could be expected by chance. Willmann (8) sent four simple abstract geometrical designs to several composers and asked them to compose appropriate musical themes after they had briefly studied each of the designs. Themes for the same designs showed marked similarities which differentiated them from other theme-design combinations. These experiments demonstrate that significant agreement of choice occurs when several *Ss* are asked to match visual designs with words, nonsense sounds, or music. The reasons for this agreement, however, remain obscure, and the choices made are largely unpredictable.

In 1952, Osgood (4) suggested that the meaning of a word might be specified by rating it on several 7-point scales, each of which has been defined by a pair of polar terms such as Kind—Cruel, High—Low, Heavy—Light. Ratings of a word on a series of such

FIG. 1. Word-sign combinations used for the study of fittingness of signs to words.

TABLE 1

SCALES USED IN SEMANTIC DIFFERENTIAL

angular—rounded	cold—hot	fast—slow
weak—strong	good—bad	hard—soft
rough—smooth	tense—relaxed	ugly—beautiful
active—passive	heavy—light	green—red
high—low	kind—cruel	sick—healthy

scales may be represented by a profile which thus becomes a kind of "profile of meaning" of that word. Because visual signs may be rated directly on this semantic differential, the degree of similarity between the connotative meanings of any given word-sign pair becomes measurable. This study was undertaken to test the hypothesis that in the situation where one of two signs is chosen as "best fitting" a word, the ratings of the word on the differential would be closer to the ratings of the chosen sign than to those of the non-chosen.

METHOD

Choice of word-sign pairs.—Ten words, each presented in connection with a pair of pictorial designs, were given to a class of 156 students in introductory psychology. The *S*s received a mimeographed sheet with the words and signs arranged as shown in Fig. 1. They were then asked to draw a circle around the letter under the sign which seemed to go best with the word.

The first six words with their signs were the same as those used by Hall and Oldfield (**3**). Of the remaining four words and eight signs, all of the words and one of the signs (Sign B for *Visionary*) were used by Hall (**2**). The remaining seven signs were drawn for this experiment to resemble Hall's more abstract designs.

The semantic differential.—In the test form for the semantic differential the 15 scales listed in Table 1 were used. These scales were selected because they are similar to those used by Osgood and because they represent the three main factors he has isolated in factor analyses of these ratings (**6**). The concepts rated were the 10 words and 20 signs shown in Fig. 1, presented in an order which did not reflect the pairings of Fig. 1. Each concept was rotated against each scale so that successive appearances of any one concept were separated by 29 other ratings.

The resultant 450-item test form was administered to 39 students who were unaware of the nature of the hypothesis and of the word-sign pairings used in Fig. 1. They were instructed to mark the scales rapidly and to judge each item as independently as possible. Ratings were given a numerical value by assigning the numbers 1 to 7 to the successive steps on each scale, the left-hand side being arbitrarily called 1. After the numerical ratings were punched on IBM cards, the cards for all the group were then put through a tabulator to obtain the distribution of ratings on each item and their sums. Mean ratings were obtained with an ordinary desk calculator. Two test forms were discarded because of unclear placement of the check marks. Thus the mean ratings are based on 37 complete forms. The difference, d, between the mean rating of a word and the mean rating of each of the two associated signs was then found on each of the 15 scales. These were combined into a single measure, D, of the distance between a given word-sign pair using the formula $D = \sqrt{\Sigma d^2}$ (**5**).

RESULTS

The results found in Table 2 confirm the significant agreement of choice found by others (**2**, **3**). The least decisive agreement was found for *Rhythm* with its associated signs, but even here tested against a hypothetical 50:50 distribution $\chi^2 = 9.82$; $P < .01$.

The results from the semantic differential given in Table 3 show that the chosen sign was closer to the word

TABLE 2

NUMBER CHOOSING SIGN A OR SIGN B FOR EACH OF THE 10 WORDS, AND THE CORRESPONDING *D* BASED ON DIFFERENCES IN MEAN RATINGS

Word	Number Choosing		*D* Based on Diff. in Mean Ratings	
	Sign A	Sign B	Sign A	Sign B
Rhythm	97	58	6.5	8.4
Wrong	143	13	4.9	6.0
Justice	34	120	5.5	4.7
Serene	128	28	5.0	10.0
Storm	138	18	5.9	7.0
Philosophy	15	141	8.2	5.3
Visionary	24	131	5.1	5.4
Crisis	5	149	11.6	3.9
Peace	156	0	3.3	11.0
Enthusiasm	4	152	8.0	7.0

than the nonchosen sign on the *D* measure in 9 of the 10 choice situations. The one exception is found in the case of *Visionary* where Sign A (the nonchosen sign) was slightly closer to the word on the *D* measure than Sign B (the chosen sign).

With 15 scales and 10 words there were 150 instances in which the mean ratings of the associated signs could be compared to that of the word. In 112 of these comparisons, disregarding the quantitative size of the difference, the mean rating of the chosen sign was closer to the word than was the mean rating of the nonchosen sign. This distribution is significantly different from what could be expected on the hypothesis that the mean rating of neither sign is more likely to be closer to the word than the other ($\chi^2 = 36.6$; $P < 0.01$).

Discussion

The results indicate that the semantic differential reflects the connotative meanings evoked both by visual designs and words. When these meanings for a word-sign pair are similar, then this combination is regarded as "fitting"; and, as seen in Table 2, marked agreement will be found in choosing such a pair over a pair with less similarity.

It is likely that the sign connotations are due to certain formal qualities which are brought out by the differential and made roughly measurable even though for most people a descriptive verbal formulation of these qualities is impossible. It is true that conventional learned-object associations may enter into these choices, although they have been fairly well excluded from many of the designs. When these conventional associations do enter, they appear to act as a separate "conventional" factor which might add to, or mask, the "formal" factor. Hall (2) has shown that, when both types of choice are possible, individuals differ considerably in the relative influence of these two factors on their choices. Since the semantic differential does not appear to be a suitable instrument for reflecting such specific associations, the relationships demonstrated in this study would likely appear most clearly when highly abstract visual designs are used.

Summary

Pairs of pictorial signs were presented in connection with each of 10 words. Statistically significant agreement of choice was found when 156 *S*s were asked to select the sign "best fitted" to each word. The same words and signs were rated independently on a semantic differential made up of 15 scales. The mean ratings of the chosen signs were found to be closer to the mean ratings of the word than were those of the nonchosen signs. It is concluded that "fittingness" of sign to word is related to the degree of similarity of their connotative meanings as indicated by ratings on the semantic differential.

References

1. Bartlett, F. C. *Remembering.* Cambridge Eng.: Cambridge Univer. Press, 1932.
2. Hall, K. R. L. The fitness of signs to words. *Brit. J. Psychol.*, 1951, **42**, 21–33.
3. Hall, K. R. L., & Oldfield, R. C. An experimental study on the fitness of signs to words. *Quart. J. exp. Psychol.*, 1950, **2**, 60–70.
4. Osgood, C. E. The nature and measurement of meaning. *Psychol. Bull.*, 1953, **49**, 197–237.
5. Osgood, C. E., & Suci, G. J. A measure of relation determined by both mean difference and profile information. *Psychol. Bull.*, 1952, **49**, 251–262.
6. Osgood, C. E., & Suci, G. J. Factor analysis of meaning. *J. exp. Psychol.*, 1955, **50**, 325–338.
7. Usnadze, D. Ein experimenteller Beitrag zum Problem der psychologischen Grundlagen der Namengebung. *Psychol. Forsch.*, 1924, **5**, 24–43.
8. Willmann, R. R. An experimental investigation of the creative process in music. *Psychol. Monogr.*, 1944, **47**, No. 1 (Whole No. 261).

(Received October 10, 1957)

Journal of Experimental Psychology
Vol. 56, No. 4, 1958

INHIBITION WITH REINFORCEMENT (CONDITIONED INHIBITION)[1]

DONALD C. KENDRICK

Institute of Psychiatry (Maudsley Hospital), University of London

Pavlov (**12**) found that conditioned stimuli while still being reinforced passed into a state of inhibition. The rate of growth of this inhibitory process was increased when the same stimulus was applied at short intervals of time, and kept to a minimal degree when the same stimulus was applied infrequently. Hovland (**8**) pointed out that there appeared to be two distinct types of experimental extinction curves: (*a*) those characterized by a continuous decline, and (*b*) those showing a larger response on the second or third extinction trial than on the first. Hovland suggested that the initial rise in the second type of curve was due to negative adaptation to continuous reinforcement. He partially verified this hypothesis. Gardner and Nissen (**6**) also observed that prolonged training often showed periods when practice seemed to have negative effects. These reports offer evidence to support the view that extinction phenomena can be observed without the withdrawal of the reinforcing agent.

Gleitman, Nachmias, and Neisser (**7**) criticized the Hullian theory of inhibition and extinction, presenting the argument that the Hullian learning equation should lead to the prediction that the ordinary learning curve should rise to a maximum and then eventually return to the base line. This prediction is deduced as follows: if a habit strength becomes asymptotic to a maximum value, reinforcement of the habit is no longer possible, and so I_R and ${}_sI_R$ should then grow unopposed with each repetition of the habit after the ${}_sH_R$ asymptote has been reached. This growth of I_R and ${}_sI_R$ will lead to progressive decrements in performance to the point where ${}_sI_R$ will equal the value of ${}_sE_R$, when the learning curve will drop to the base line never to recover again. They add that this extinction should be achieved in the same number of trials required for experimental extinction. Calvin, Clifford, Clifford, Bolden, and Harvey (**3**) reported an experiment in which trained rats were to run a straight runway for a food reinforcement. After a period of consecutive testing of 30 trials per day per *S*, it was found that all *S*s refused to run for their food after a few trials, although they were still deprived of food and received a reward on every trial. The procedure in this experiment was criticized by Dinsmoor (**4**) who pointed out several reasons, other than the Gleitman prediction, why the *S*s may have stopped running, the most pertinent of which was that Calvin may have inadvertently trained the *S*s not to run, because when any *S* refused to run it was picked up and placed in the goal box with the reward present, thus rewarding the nonrunning behavior.

It was therefore decided to repeat Calvin's experiment in a modified form, with more stringent nonrunning criteria exercised.

[1] This work was carried out with the aid of a grant from the Bethlem Royal and Maudsley Hospital Research Fund.

METHOD

Subjects.—The *Ss* were 10 male albino rats which formed part of the sixth generation in a selective breeding study of emotionality (2). For reasons not pertinent to this report the 10 *Ss* were composed of two groups each of five brothers. Group A was the offspring of a fifth generation pair of emotional *Ss*, and Group B was the offspring of a fifth generation pair of non-emotional *Ss*. The *Ss* were all experimentally naive and exactly 100 days old when testing began. Their experience prior to testing was rigidly controlled; one month before testing each *S* was housed in a separate cage in the animal room, and was kept under the same conditions as far as temperature, light/dark cycle, diet, and husbandry were concerned. Two weeks before testing, each *S* was handled for 5 min. every day and gradually brought onto a 23-hr. water deprivation schedule. During testing careful check was kept on each *S*'s weight. No loss in weight was recorded, and no *S* suffered from any detectable illness.

Apparatus.—Two identical open-top runways in prepared wood were constructed, 12 ft. long and 6 in. wide, and with sides 10 in. high. The first foot of the runway was made into a starting box by fixing a door 1 ft. from the end. The door was opened by a spring mechanism, allowing it to swing from the vertical position to the horizontal, so that *S* had to go underneath the door to get into the runway. A burette stand with a 6 × 12-in. base was placed at the opposite end. A 100-cc capacity burette was fitted in the stand and a watchglass 1¼ in. in diameter was fixed directly beneath the nozzle of the burette, the tip of the nozzle being 3 in. above the watchglass. One-fourth cubic centimeter of water could be accurately measured and dropped into the glass without splashing. Both runways were fitted to the opposite walls of a 5 × 14 ft. room, 2.5 ft. from the floor, and were painted flat black. The room was illuminated by a 100-w. pearl bulb, which was placed at a point midway between the two runways and halfway along each runway. The room temperature during testing was kept within a few degrees of 72° F.

Each runway was fitted with two clocks measuring accurately to .01 sec. and operated by a capacity change relay system. One clock measured latency, or the interval between the opening of the door and the *S* moving under the door, and the second clock measured speed of running time, or the interval from the *S* passing the door to reaching the burette stand.

Procedure.—The day before testing began each *S* was weighed and the total water intake over the hour's drinking period was measured for each *S*; it was found that Group A's mean weight did not differ significantly from Group B's ($t = .40$ with 8 *df*) nor did the two groups' water intake differ significantly ($t = 1.58$ with 8 *df*). Each *S* from Group A was paired with an *S* from Group B on the basis of rank order in weight, the heaviest pair of *Ss* always being tested first each day, and the other pairs followed in descending order of weight.

Each pair of *Ss* was run at the same time, their allocation to a particular runway on a given day being determined by a Gellerman series, so that in 10 days' testing each *S* ran five days in each runway, but did not run in the same runway for more than three consecutive days. Each *S* of each pair mirror-imaged its partner's allocation to runways. All *Ss* learned to run the runways in three days of free exploration with reward present in the dish.

The testing procedure was as follows: each *S* was placed in the starting box facing the door, the door was opened and *S* ran to the reinforcement dish. Immediately *S* finished drinking, it was taken out of the goal box and run again. This was repeated for 30 trials for each *S*. The reinforcement per trial was ¼ cc, so that if *S* completed the 30 trials it received 7.5 cc. of water. Three hours after the last trial of each *S* a further drinking period ad libitum of ½ hr. was allowed. The *Ss* were thus tested under approximately 20 hr. water deprivation.

The criterion for extinction was that on three consecutive days *S* must refuse to run for water twice consecutively within the first five trials. A refusal was defined as *S*'s failing to stop the running time clock within 5 min. of the door being opened. After the first refusal *S* was momentarily taken out of the starting box and then put back in again and the door opened. If any *S* had two consecutive refusals during any part of a day's testing, testing was discontinued for that day, but was not counted as an extinction day. On no occasion was *S* lifted out of the runway and placed in the goal box for reinforcement if it refused to run. On every trial it was ensured that *S* had drunk the water and that the paws and fur were relatively dry before starting another trial.

Thirty-six days after the last extinction day, each *S* was re-run under the same conditions to see if there was any spontaneous recovery of the habit.

RESULTS

The main result of the experiment was that all *Ss* met the criterion for nonrunning. Seven *Ss* ran zero trials over the three extinction days. The mean number of days for Groups A

and B to reach the nonrunning criterion was 32.40 and 34.00, respectively; the difference between the groups is not significant ($t = .35$ with 8 *df*). Although two measures were recorded, only the speed of running data will be discussed as it was found that Group A's latencies had a very large variance and could not be normalized by various transformations to be used reliably in statistical tests of significance between means which are being reported for the purpose of this paper. Group B's latencies followed very closely the trend of their speed of running results.

Because there were considerable differences in the number of days between *S*s in reaching the criterion for nonrunning, the production of composite learning curves has been effected by breaking the individual *S*'s learning curve into two parts. The results of 12 days' testing from each *S* has been utilized in the following way: as the first three days were training days, these have been omitted and Day 4 to 9 have been designated the learning period, or Period 1. These days in point of time were the same for each *S*. Then taking the first extinction day for each *S*, the six days immediately preceding this point have been designated the extinction period, or Period 2. Thus the latter part of each *S*'s learning curve has been equated back from the nonrunning point, but these days are not equated in time. The range of days omitted by this procedure is from 1 to 23. The median running time of each *S* was computed for the 12 days, then the group mean median speed of running time was calculated for each day and plotted as shown in Fig. 1. It was found that there was no difference between the groups in Period 1 (mean running time Group A = 10.56 sec.; Group B = 8.71 sec., $t = 1.005$ with 58 *df*), but that Group A *S*s were running significantly slower than Group B in Period 2 (mean running time Group A = 10.52 sec., Group B = 4.85 sec.; $t = 3.83$ with 58 *df*, $P = .01$). The plotted curves show no evidence for a progressive gradual slowing down in speed of running.

It was noted on looking at the running times over a day that there appeared to be certain quantitative differences between the groups which were being masked by taking only the median running times. Group A after a period of time was showing a distinct "warm-up" effect. The first few trials per day were slow,

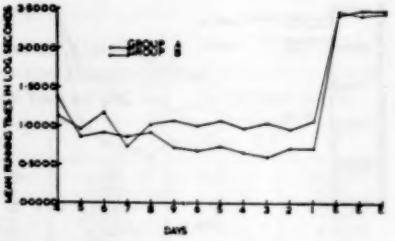

FIG. 1. Performance curves based on the mean of the five median running times per day per group.

gradually increasing to a peak which was maintained for several trials, then the speeds began to decrease again, often slowing to such an extent as to produce two consecutive 5-min. criterion periods before Trial 30 had been reached. The *S*s of Group B by observation were showing none of these effects until a few days before extinction and then only on a very small scale. To test this observation each *S*'s daily running times were re-worked in the following way. The total trials per day per *S* (excluding any 5-min. periods, as these are not counted as a trial because there is no reinforcement of the running habit by drinking water) were divided into fifths and a mean taken of each fifth. These points were plotted on a graph and it was found that there was a very pronounced trend exhibited by the curves of the first and last fifth sections of Group A's speed of running times. These curves showed a gradual rise to a peak of performance followed by a gradual decline towards the base line (5 min.). The curves based on the middle fifths of Group A showed also a gradual rise to a peak followed by a plateau inclining minimally towards the base line. Group B showed none of these effects until a few days before they refused to run, and then the slowing down of the first fifth of the times was very small compared with that of Group A.

An analysis of variance design was set up in order to test these observations. To keep the population within each cell constant, it was decided to test the differences in running speed on the means of the first and last six running times per day per *S*, in Periods 1 and 2, irrespective of the total completed trials on these days. The curves based upon the analyses of

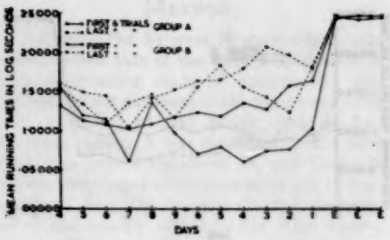

FIG. 2. Performance curves based on the mean of the means of the first and last six trials per *S* per day.

variance data is shown in Fig. 2 and the results of the analyses are shown in Table 1.

The mean of the first and last six trials allowed the maximum number of running times within each cell without any time score being used twice, because prior to extinction the smallest number of trials recorded on any day was 12. The scores were converted to logarithms to normalize the data. In the analysis of both sets of data the between-rat, within-strains effect is highly significant, which means that any conclusions drawn from these analyses must always be tentative. In the analysis of the first six trials the *F* ratio between *S* × Period/Period × Strain is 31.409, which is significant at the .001 level, indicating that although each *S* acts individually, there is still a considerable strain difference in the manner in which the nonrunning effect is produced. In the analysis of the last six trials there are no group differences over Periods, but both groups are running significantly slower in Period 2, showing that there is a shift towards the base line in this section.

The observation that Group A appeared to run 30 trials less often than Group B was also tested statistically. The proportion of times that each group ran 30 trials per day was calculated: for Group A it was .44 and for Group B it was .79. The arc sine transformation of the square roots of the proportions was then taken and a critical ratio set up which yielded a value of 6.091 which is significant above the .01 level, showing that Group A did complete the full 30 trials on fewer occasions than Group B.

No spontaneous recovery was recorded for any *S* in the retest 36 days after extinction was complete.

DISCUSSION

The fact that there was no spontaneous recovery of the habit suggests that the mechanism by which this nonrunning behavior was produced was other than experimental extinction. Gleitman et al. (7) were correct in their prediction that the repetition of a habit would produce the extinction of that habit, but their theoretical argument is at fault because Hull (9) was very inconsistent with his theory of inhibition. Hull's principal formula of ${}_{s}E_{R}$ formation, ${}_{s}E_{R} = {}_{s}H_{R} \times D$, makes it inconsistent to then add the negative drive I_R with the negative habit ${}_{s}I_{R}$. Iwahara (10) has discussed this inconsistency and suggested that I_R and ${}_{s}I_{R}$ should be multiplicative. He finds this multiplication contradictory because if I_R was zero, then the total inhibitory strength must also be zero. This obviously contradicts the facts. Jones (11) has produced another formula which allows ${}_{s}I_{R}$ to be activated without the presence of I_R. His formula is: ${}_{s}\bar{E}_{R} = ({}_{s}H_{R} - {}_{s}I_{R}) \times (D - I_{R})$, which multiplied out gives: ${}_{s}\bar{E}_{R} = (D \times {}_{s}H_{R}) - (D \times {}_{s}I_{R}) - ({}_{s}H_{R} \times I_{R}) + ({}_{s}I_{R} \times I_{R})$. In this formula ${}_{s}I_{R}$ can be activated by D as well as I_R. It was found in this experiment that seven *S*s refused to run consistently on the first two trials. If I_R and ${}_{s}I_{R}$ were simply multiplicative this result would be impossible, as after a considerable rest period presumably all I_R would have dissipated and there would be no drive activating ${}_{s}I_{R}$. But Jones' formula allows both the positive

TABLE 1

SUMMARY OF ANALYSES OF VARIANCE

Source	*df*	First 6 Trials per Day		Last 6 Trials per Day	
		MS	*F*	*MS*	F
Periods (P)	1	.2410	2.87	3.2912	26.44**
Days within Periods	10	.2294	2.73*	.1223	0.98
Strains (St.)	1	2.2422	26.72**	.9705	7.80**
Ss within Strains	8	.5792	6.90**	1.1085	8.90**
P × St.	1	3.2760	39.05**	.2018	1.62
Ss × Period	8	.1043	1.24	.7450	5.98**
Day × St.	10	.0124	0.15	.1924	1.55
Residual	80	.0839		.1245	
Total	119				

* $P = .01$.
** $P = .001$.

and negative habits to be aroused by the positive primary drive (D). Therefore if ${}_sH_R = {}_sI_R$ after all I_R has dissipated, there will still be nonrunning behavior because both habits are multiplied by the same strength of the primary drive. It is therefore suggested that the original learning equation cannot be accepted because it is illogical, and that the reformulation by Jones is in some measure supported by certain results in this experiment.

The fact that the learning curves based upon the median running times do not show a gradual return to the base line can be explained in the following way. Eysenck (5) has suggested that warm-up effects found under massed practice on the pursuit-rotor are not due to the warm-up decrement as proposed by Ammons (1), but that ${}_sI_R$, being a habit, must react to nonreinforcement like any other habit and undergo extinction. Thus, after a rest period during which I_R has dissipated, on the first few trials little or no I_R is accumulated and so not dissipated. Consequently ${}_sI_R$ is not reinforced and so undergoes extinction up to the point where once again I_R is dissipated and reinforces the ${}_sI_R$. If this system is to be taken to its logical conclusion, then if ${}_sI_R$ undergoes experimental extinction it must also spontaneously recover, allowing the strength of ${}_sI_R$ to increase from period to period. It is thought that in the case of this experiment this process occurred, the ${}_sI_R$ growing until it equaled the strength of ${}_sH_R$ or even becoming stronger as shown by the fact that so many consecutive refusals were recorded. If ${}_sI_R$ only equals ${}_sH_R$ in strength, then there must be equal chances of either event occurring, but this was not the case with *Ss* in this experiment. Because ${}_sI_R$ extinguishes during the first few trials, the middle section of the trials keeps a relatively constant level, the median time generally falling within this section, and so there was no evidence for a progressive decrease in running times shown by a graph based upon the mean median speed of running times.

Group A appears to show the production of ${}_sI_R$ more clearly than Group B. The results have shown that (*a*) by measurement of water intake the groups do not differ in drive levels, (*b*) in Period 1 there are no differences between the groups in speed of running, these two facts suggesting that the groups are well matched at this point, and (*c*) the major differences appear to be in the warm-up effects and in the number of trials completed each day per group. It is thought that these differences can be accounted for by postulating both individual and group differences in the strength of ${}_sI_R$ needed before the habit reaches a superthreshold level. In the case of Group A

the threshold is low, allowing the superthreshold growth of $_{S}I_{R}$ to be observed over a considerable period by its depressing effect on running times at the beginning of a testing session, whereas the slowing down towards the end of the session is due to the combined effects of I_R and $_{S}I_{R}$. In Group B the threshold is very much higher, almost as high as the strength of $_{S}H_{R}$ itself, so that very little evidence of $_{S}I_{R}$ is shown until its strength is almost equal to that of $_{S}H_{R}$; hence, the rather abrupt nonrunning behavior in Group B.

It is therefore considered that Gleitman's main prediction has been verified, in that the constant repetition of a habit leads to its extinction, but that the mechanisms underlying this extinction are other than those proposed by Gleitman.

Summary

Ten albino rats were trained to run a 10-ft. runway. On each trial they received ½ cc of water as reward, being tested after 20 hr. water deprivation. The *S*s ran 30 massed trials each day. After a period of days all *S*s refused to run for their water. The results have been explained in terms of a reformulation of the Hullian learning equation, of the extinction and recovery of conditioned inhibition, and differences in the threshold level of conditioned inhibition.

References

1. Ammons, R. B. Acquisition of a motor skill: I. Quantitative analysis and theoretical formulation. *Psychol. Rev.*, 1947, **54**, 263–281.
2. Broadhurst, P. L. Studies in psychogenetics: The quantitative inheritance of behaviour in rats investigated by selective and cross-breeding. *Bull. Brit. psychol. Soc.*, 1958, **34** 2A. (Abstract)
3. Calvin, A. D., Clifford, L. T., Clifford, B., Bolden, L., & Harvey, J. An experimental validation of conditioned inhibition. *Psychol. Rep.*, 1956, **2**, 51–56.
4. Dinsmoor, J. A. Absurdum revisited; A comment. *Psychol. Rep.*, 1956, **2**, 255–256.
5. Eysenck, H. J. "Warm-up" in pursuit rotor learning as a function of the extinction of conditioned inhibition. *Acta Psychol.*, 1956, **12**, 349–370.
6. Gardner, L. P., & Nissen, H. W. Simple discrimination behavior of young chimpanzees: Comparisons with human aments and domestic animals. *J. genet. Psychol.*, 1948, **72**, 145–164.
7. Gleitman, H., Nachmias, J., & Neisser, U. The S-R reinforcement theory of extinction. *Psychol. Rev.*, 1954, **61**, 23–33.
8. Hovland, C. I. "Inhibition of reinforcement" and phenomena of experimental extinction. *Proc. nat. Acad. Sci.*, 1936, **22**, 430–433.
9. Hull, C. L. *Principles of behavior*. New York: D. Appleton-Century, 1943.
10. Iwahara, S. Hull's concept of inhibition: A revision. *Psychol. Rep.*, 1957, **3**, 9–10.
11. Jones, H. G. The status of inhibition in Hull's system: A theoretical revision. *Psychol. Rev.*, 1958, **65**, 179–182.
12. Pavlov, I. P. *Conditioned reflexes*. London: Oxford Univer. Press, 1927.

(Received October 14, 1957)

Journal of Experimental Psychology
Vol. 56, No. 4, 1958

EXTINCTION AND SPONTANEOUS RECOVERY OF CONDITIONED EYELID RESPONSES AS A FUNCTION OF AMOUNT OF ACQUISITION AND EXTINCTION TRAINING [1]

WILLIAM F. PROKASY

Pennsylvania State University

To date, little attention has been paid to the variables of amount of acquisition and extinction training as they affect responding during extinction and spontaneous recovery test trials. The lack of information is more notable in the area of classical conditioning. Pavlov (**13,** pp. 58, 59) suggests that the greater the amount of extinction training, the less the amount of responding during spontaneous recovery test trials. In the only pertinent classical conditioning studies reported in the American literature, Kimble and Dufort (**9**) and Hovland (**8**) report greater resistance to extinction with a greater number of acquisition trials.

The relevant literature in instrumental conditioning is difficult to assess because of gross methodological differences among the studies. When response latency in a straight runway is employed as the measure, whether or not more acquisition training leads to higher resistance to extinction depends upon the discriminability of the start and goal boxes (**3, 10, 11, 12**).

The several studies (**1, 7, 14, 19, 20**) concerned with rats and their lever-pressing behavior report essentially the same effect of differing amounts of acquisition training: as acquisition training increases, resistance to extinction increases. Williams (**19**) and Homme (**7**) also report that resistance to extinction during the spontaneous recovery test trials is an increasing function of the amount of acquisition training. This, however, is but a tentative finding, since in both studies initial extinction training came 24 hr. after the end of acquisition training.

The object of the present study was to investigate the effects of amount of both acquisition and extinction training upon extinction and spontaneous recovery. The asset of the particular design employed is that it enables simultaneous assessment of both the acquisition and extinction variables upon spontaneous recovery.

METHOD

Apparatus.—The Wisconsin eyelid-conditioning apparatus, described elsewhere (**6, 16**), was employed. Electronic differentiators permitted the recording of both the response and its first time derivative. The derivative permitted a more accurate recording of response latencies.

The CS lasted for 500 msec., and the UCS was administered 500 msec. after the onset of CS. The intertrial interval ranged from 25 to 35 sec., with an average of 30 sec. CR's were defined as any response that occurred between 200 and 500 msec. after the onset of the CS. The UCS was a puff of air to the left cornea, generated by the fall of a 150-mm. column of mercury. The CS was a sudden change in brightness of a milk-glass window, about 80 cm. in front of *S*, from 7 to 100 mL.

[1] The results reported constitute a portion of a dissertation submitted to the Graduate School of the University of Wisconsin in partial fulfillment of the requirements for the Ph.D. degree. The author is indebted to his advisor, David A. Grant, as well as to F. A. Mote and L. B. Wyckoff, for his support and constructive criticism which contributed to the completion of the reported research.

Subjects.—The *Ss* were 92 men and women volunteers from introductory psychology courses at the University of Wisconsin. They were assigned randomly to one of the 18 experimental conditions, and received partial course credit for their services. Two *Ss* were discarded because they made no recordable UCR to the puff stimulation, so that the data of only 90 *Ss* were employed in the final analyses.

Design and procedure.—The experiment involved a 3 × 3 orthogonal design with three levels each of acquisition and extinction training. The *Ss* received 20, 40, or 60 acquisition trials followed by 0, 10, or 30 extinction trials. Approximately 24 hr. later each *S* was given 30 spontaneous recovery test trials. All *Ss* received neutral instructions and were run under light-adapted conditions in order to minimize the presence of sensitized beta responses (4).

Results and Discussion

Extinction.—The percentage of CR's in blocks of five extinction trials during extinction is presented in Fig. 1, and analyses of variance and trend are summarized in Table 1.[2] Each of the points for the first 10 trials are based upon 20 *Ss*. The analyses of variance and trend are based upon the 10 *Ss* in each of the three acquisition groups who received all 30 extinction trials.

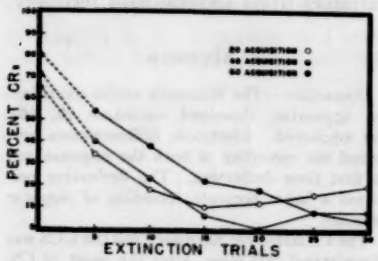

Fig. 1. Percentage of CR's during extinction in five-trial blocks of the T groups. The initial point on each curve is the average percentage of CR's for the final five acquisition trials for each group.

[2] An arcsine transformation is suggested by Snedecor (18), and is used in all subsequent analyses of percentage of CR scores.

TABLE 1

Analysis of Variance and Trend of Extinction Data

Source	Error Term	*df*	*F*
A. Over-all trend	E	5	5.98**
1. Linear	E-1	1	11.50**
2. Quadratic	E-2	1	1.30
B. Between-group means	D	2	2.90
C. Between-group means	E	10	3.28**
1. Linear	E-1	2	3.66*
2. Quadratic	E-2	2	6.05**
D. Between-indiv. means	E	27	3.36**
E. Between-indiv. trend		135	(344.62)
1. Linear		27	(804.82)
2. Quadratic		27	(371.15)

Note.—Trend components beyond the quadratic were analyzed, but are omitted here since none proved to be significant. Values in parentheses are mean squares.

* $P = .05$.
** $P = .01$.

As shown in Table 1, the over-all linear trend is significant, which implies that the best fitting straight line to all points deviates significantly from slope zero. Due to the asymptoting nature of extinction curves, a significant over-all quadratic component was also expected, but failed to materialize.

The significant between-groups linear and quadratic components of trend, shown in Table 1, in conjunction with the nearly parallel extinction curves of the 40- and 60-trial acquisition groups suggest a stable upturn in the extinction curve of the 20-trial acquisition group. Such upturns have occurred in the data of Reynolds (17) with massed extinction training, by Grant, Riopelle, and Hake (5) with patterns of reinforcement, and by Grant, Schipper and Ross (6) with intermittent reinforcement and massed training. While it is possible that these upturns are an artifact of a few unusual

Ss, the frequency with which they show up suggests that an investigation of the phenomenon is in order.

Spontaneous recovery.—Average percentage of CR's in the spontaneous recovery test is shown in Fig. 2 in blocks of five trials. Table 2 summarizes the analyses of variance and trend based upon the five-trial blocks. Although some groups show reversal, the over-all linear trend is significant, indicating an over-all decrease in responding during the spontaneous recovery test trials. As in the case of the extinction data, the expected significant over-all quadratic trend did not occur.

Between-group mean percentages of CR's as a function of both acquisition and extinction are significant.

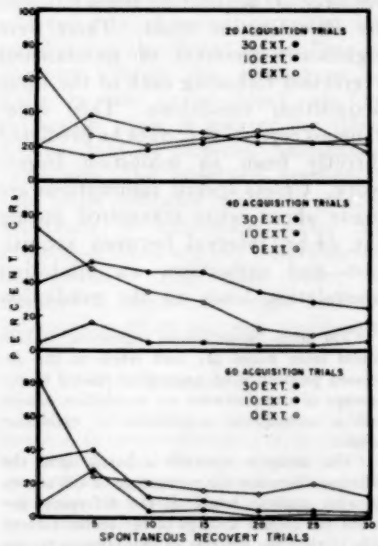

FIG. 2. Percentage of CR's in spontaneous recovery in five trial blocks as a function of amount of acquisition and extinction training. Number of extinction trials is the parameter. The initial point on each curve is the total percentage of CR's on the last five acquisition or extinction trials for each group.

TABLE 2

ANALYSIS OF VARIANCE AND TREND OF SPONTANEOUS RECOVERY DATA

Source	Error Term	*df*	*F*
A. Over-all trend	(E)	(5)	[6.64**]
1. Linear	E-1	1	15.97**
2. Quadratic	E-2	1	2.23
B. Between-group means	(D)	(8)	[3.18**]
1. Acquisition	D	2	4.89**
2. Extinction	D	2	4.73*
3. Interaction	D	4	1.54
C. Between-group trends	(E)	(40)	[1.50*]
1. Linear	(E-1)	(8)	[1.90]
a. Acquisition	E-1	2	2.76
b. Extinction	E-1	2	—
c. Interaction	E-1	4	1.94
2. Quadratic	(E-2)	(8)	[2.44*]
a. Acquisition	E-2	2	3.18*
b. Extinction	E-2	2	—
c. Interaction	E-2	4	3.16*
D. Between-indiv. means	E	81	3.93**
E. Between-indiv. trends		405	(371.88)
1. Linear		81	(630.03)
2. Quadratic		81	(338.07)

Note.—Trend components beyond the quadratic were analyzed, but are omitted here since none proved to be significant. Values in parentheses are mean squares.
* $P = .05$.
** $P = .01$.

The total percentages of CR's during the spontaneous recovery test trials are 25.4, 22.5, and 11.7 following 20, 40, and 60 acquisition trials, respectively. The direction of this relationship was entirely unexpected, and does not conform to the sparse data (**7, 19**) which previous investigators have reported. These data are also opposite in direction to what would be predicted from statistical learning theory as presented by Homme (**7**); and they do not appear to be predictable from any theory which views extinction as the process of inhibition overcoming an established S-R relationship. More parametric informa-

TABLE 3

t TESTS OF SPONTANEOUS RECOVERY AND REGRESSION

Extinction Trials	Acquisition Trials		
	20	40	60
0	4.49**	2.39*	5.81**
10	−1.64	−.68	−.80
30	.43	−1.62	−3.85**

* $P = .05$.
** $P = .01$.

tion on this result is desirable, since it is most likely a result of the particular range of acquisition trials employed in this study. More explicitly, for example, with fewer than 20 acquisition trials it might be expected that response strength would not have built up enough to permit much, if any, spontaneous recovery.

As a function of 0, 10, and 30 extinction trials, total percentage of CR's in spontaneous recovery test trials are 23.7, 25.5, and 10.7. In general, and in accord with prior research (**1, 7, 8, 14, 15, 19, 20**), the significant between-groups means as a function of amount of extinction training indicates that as extinction training increases, the amount of responding in spontaneous recovery decreases.

Differences in the quadratic components of trend due to the variable of amount of acquisition training were significant. This appears to be primarily a function of the contrast of the spontaneous recovery curves of the 20-trials group with the 40- and 60-trials groups. In Fig. 2, two of the three curves following 20 acquisition trials, if anything, suggest an upturn, while the curves following the other acquisition levels are of essentially downward trend.

The quadratic components of trend related to the interaction of acquisition and extinction training is significant. This trend component can become significant in such cases as, for example, when there is no difference in the quadratic components across all extinction levels for a particular acquisition level, while a difference does exist across extinction levels for another acquisition level. The meaning of the interaction in the present case, however, is obscure.

Table 3 presents *t* tests of spontaneous recovery and regression.[3] The tests are based upon the differences between the percentage of CR's in the final five acquisition or extinction trials and the first five spontaneous recovery test trials.[4] The only significant amount of spontaneous recovery, as indicated by this measure, was shown by the group that had received 60 acquisition trials followed by 30 extinction trials. There were significant amounts of spontaneous regression following each of the three acquisition conditions. This spontaneous regression cannot be predicted directly from an inhibition framework. Unless special assumptions are made about what transpired during the 24-hr. interval between acquisition and extinction, an inhibition formulation leads to the prediction

[3] The term "spontaneous regression" is borrowed from Estes (2), and refers to the decreased proportion of responding related to the passage of time between an acquisition session and a subsequent acquisition or extinction session.

[4] The analysis reported is based upon the difference between the percentage of CR scores. Another analysis based on the differences between the arcsine analogs of the scores differed only in that the amount of spontaneous regression following 40 acquisition trials fell short of the .05 level of significance. Note, too, that employing percentage of CR's over the last five trials of acquisition or extinction and the first five spontaneous recovery trials introduces a bias with respect to an "ideal" measure taken on the first spontaneous recovery trial.

TABLE 4

PERCENTAGE OF EXTINCTION CR's

Extinction	Number of Acquisition Trials		
	20	40	60
Immediate	16.3	8.7	25.7
Delayed	25.7	24.3	17.7

of a certain amount of above-acquisition-level performance at the beginning of the extinction session.

Table 4 presents total percentage of CR's for those groups of *Ss* who received 30 extinction trials immediately and those who received 30 extinction trials 24 hr. after the end of acquisition training. Although total responses are, unexpectedly, greater in general across all acquisition conditions for the groups which received delayed extinction trials, this trend is not significant by analysis of variance. Since the percentage of CR's in the initial block of five extinction trials is greater for those groups which received immediate extinction, the observed nonsignificant trend indicates relatively greater resistance to extinction when the extinction trials are not administered immediately. The main effect of number of acquisition trials and the interaction effect are not significant.

SUMMARY

Ninety introductory psychology students served in an eyelid-conditioning study designed to determine the effects, on extinction and spontaneous recovery, of varied amounts of acquisition and extinction training. It was found that with an increase in acquisition training resistance to extinction increased while amount of responding during spontaneous recovery test trials decreased.

In accord with prior research, an increase in the amount of extinction training decreased resistance to extinction during spontaneous recovery test trials. The significant amounts of spontaneous regression were interpreted as not favorable for an inhibition theory of extinction, while the high rate of spontaneous recovery responding associated with the fewest number of acquisition trials was considered to be inconsistent with statistical learning theory.

REFERENCES

1. ELLSON, D. G. Quantitative studies of the interaction of simple habits: I. Recovery from specific and generalized effects of extinction. *J. exp. Psychol.*, 1938, **23,** 339–358.
2. ESTES, W. K. Statistical theory of spontaneous recovery and regression. *Psychol. Rev.*, 1955, **62,** 145–154.
3. FINGER, F. W. Retention and subsequent extinction of a simple running response following varying conditions of reinforcement. *J. exp. Psychol.*, 1942, **32,** 120–133.
4. GRANT, D. A., & NORRIS, E. B. Eyelid conditioning as influenced by the presence of sensitized beta-responses. *J. exp. Psychol.*, 1947, **37,** 423–433.
5. GRANT, D. A., RIOPELLE, A. J., & HAKE, H. W. Resistance to extinction and the pattern of reinforcement and the conditioned eyelid response. *J. exp. Psychol.*, 1950, **40,** 53–60.
6. GRANT, D. A., SCHIPPER, L. M., & ROSS, B. M. Effects of intertrial interval during acquisition on extinction of the conditioned eyelid response following partial reinforcement. *J. exp. Psychol.*, 1952, **44,** 203–210.
7. HOMME, L. E. Spontaneous recovery and statistical learning theory. *J. exp. Psychol.*, 1956, **51,** 205–212.
8. HOVLAND, C. I. 'Inhibition of reinforcement' and phenomena of experimental extinction. *Proc. Nat. Acad. Sci.*, 1936, **22,** 403–433.
9. KIMBLE, G. A., & DUFORT, R. H. The associative factor in eyelid conditioning. *J. exp. Psychol.*, 1956, **52,** 386–391.
10. LAWRENCE D. H., & MILLER, N. E. A possible relationship between reinforcement and resistance to extinction produced by removing a source of confusion from a technique that had produced opposite results. *J. exp. Psychol.*, 1947, **37,** 494–509.
11. MOTE, F. A. The effect of different amounts of reinforcement upon the acquisition and extinction of a simple running response. *J. exp. Psychol.*, 1944, **34,** 216–226.
12. MOTE, F. A., & FINGER, F. W. The retention of a simple running response after

varying amounts of reinforcement. *J. exp. Psychol.*, 1943, **33,** 317–322.

13. Pavlov, I. P. *Conditioned reflexes* (translated by W. H. Gantt). New York: International, 1928.
14. Perin, C. T. Behavior potentiality as a joint function of the amount of training and degree of hunger at the time of extinction. *J. exp. Psychol.*, 1942, **30,** 93–113.
15. Prokasy, W. F. Extinction and spontaneous recovery of conditioned eyelid responses as a function of varying amounts of acquisition and extinction training. Unpublished Ph.D. dissertation, Univer. of Wisconsin, 1957.
16. Prokasy, W. F., Grant, D. A., & Myers, N. A. Eyelid conditioning as a function of UCS intensity and intertrial interval. *J. exp. Psychol.*, 1958, **55,** 242–246.
17. Reynolds, B. Extinction of trace conditioned responses as a function of the spacing of trials during acquisition and extinction series. *J. exp. Psychol.*, 1945, **35,** 81–95.
18. Snedecor, G. W. *Statistical methods.* (4th ed.) Ames, Iowa: Collegiate Press, 1946.
19. Williams, S. B. Resistance to extinction as a function of the number of reinforcements. *J. exp. Psychol.*, 1938, **23,** 506–522.
20. Youtz, R. E. P. The change with time of a Thorndikean response in the rat. *J. exp. Psychol.*, 1938, **23,** 128–140.

(Received November 26, 1957)

Journal of Experimental Psychology
Vol. 56, No. 4, 1958

PURSUIT ROTOR PERFORMANCE AS A FUNCTION OF DELAY OF INFORMATION FEEDBACK [1]

E. JAMES ARCHER AND GEDIMINAS A. NAMIKAS

University of Wisconsin

Reynolds and Adams (3) have reported a study in which delay of information feedback produced improved performance during learning and sustained superior performance during a nonreinforced transfer condition. More recently Archer, Kent, and Mote (1) reported the results of a long-term learning experiment in which delay of information feedback was studied also. Unlike the Reynolds and Adams study, which used the rotary pursuit task, the latter study used a more complex tracking task, the Pedestal Sight Manipulation Test (PSMT). Also unlike the Reynolds and Adams study, Archer, Kent, and Mote obtained negative results in terms of effect of delay of information feedback.

The present study was intended as a retest of the Reynolds and Adams' phenomenon. Two changes were made in the conditions, however: (*a*) the number of transfer trials was increased to better observe any changes in the anticipated superior performance, and (*b*) after the specified delay the feedback to *S* was continuous rather than discrete. It was anticipated that the latter change would maximize the facilitating effect of delayed information feedback.

[1] This research was supported by a grant from the Research Committee from funds provided by the Wisconsin Alumni Research Foundation. The writers are also indebted to the Numerical Analysis Laboratory for its assistance in the statistical analyses involved in this research.

METHOD

Subjects.—Seventy-five men students from introductory psychology courses at the University of Wisconsin served individually for one session. The men were assigned randomly to one of five groups of 15 *S*s each.

Feedback conditions.—The five groups corresponded to the duration *S* had to be continuously on target before he heard a 1000-cps tone. These delay of information feedback intervals were .0, .2, .4, .8, and 1.6 sec. The tone signal was produced by a Jackson audio oscillator and a 5-in. loudspeaker. The speaker was mounted about 1 ft. from the pursuit rotor and produced a clearly audible but not unpleasant tone. The tone was heard continuously for the time *S* remained on target in excess of his group's delay condition. If, for example, an *S* in the .2-sec. group made contact with the target for .5 sec. continuously, he heard a tone for the last .3 sec.

Apparatus.—The *S*s tracked a .75-in. brass target which revolved at 60 rpm on a 5-in. orbit. The turntable had a radius of 6 in. and its speed of rotation was calibrated with a stroboscope before each *S* began his session. The *S* used a hinged stylus which had a 1/16-in. rolling steel ball for a tip.

The *S*'s performance was recorded by a Tracking-time Analyzer (1). Since this device recorded a frequency distribution of various durations of time continuously on target, it also was used to provide *S* with the delayed feedback signal. Whenever the counter corresponding to the delay period was fired, *S* heard a 1000-cps tone.

In order to restrict the feedback to the tone, the Tracking-time Analyzer was enclosed in a partially soundproof box. In addition, a large electric fan was turned on to mask any residual cues. Under these conditions the tone was the only information feedback available to *S* in addition to his own visual cues. The analyzer was calibrated twice a day, before the morning and afternoon testing sessions.

Instructions.—All *S*s were given general instructions regarding the nature of the task, including a demonstration of the best way to track the target, and of the tone. Each *S* was told the length of time he had to stay on target continuously in order to hear the tone and the

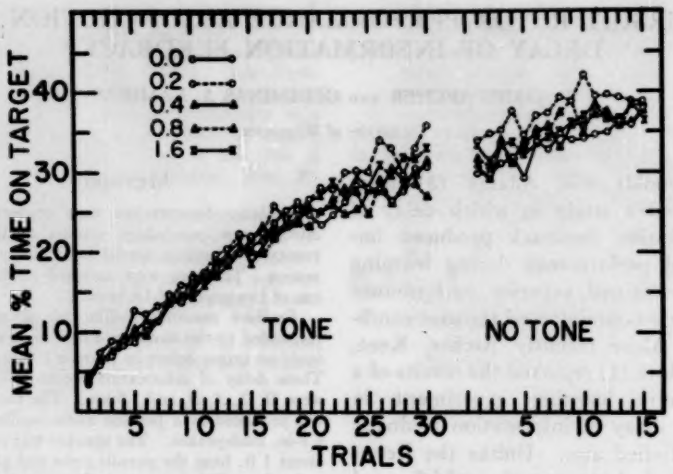

FIG. 1. Mean percentage of time on target as a function of trials of practice. The tone feedback could only be heard during the first 30 trials. The delay, in seconds, of information feedback for the five groups is indicated by the values in the upper left corner.

length of a trial and intertrial rest, but he was not told the number of trials or the specific length of the experimental session. The *S*s were told to look at magazines during rest intervals. Answering any questions except those specifically related to instructions was deferred until the end of the experimental session. Instructions were tape recorded and were identical except when different delay intervals were indicated.

Task.—All *S*s had forty-five 30-sec. practice trials with a 30-sec. rest between each trial. Only during the first 30 trials could *S* receive the tone feedback. After Trial 30, and during the regular 30-sec. rest, the loudspeaker was bypassed and *S* did not hear the tone again no matter how long he stayed on target. The *S* was informed of the change of conditions at this time but was encouraged to make the best possible score on every trial.

RESULTS

In addition to the total time on target (TOT) per trial, the longest hit (LH) per trial and the mean hit (MH) per trial were determined for each *S*. The first of these was obtained from the calibrating clock of the Tracking-time Analyzer. The second was taken as the midpoint of the interval in excess of the highest counter activated. The last named measure was determined by dividing the total time on target by the total number of hits.

As would be expected, these three measures are highly correlated with each other. During Trials 1–30, in which the tone was being presented, TOT correlated .95 with MH and .89 with LH. The correlation between MH and LH was .87 ($\sigma_r = .02$). During the last 15 trials (31–45), when no tone was presented, the correlation of TOT with MH was .96 and with LH it was .81. The correlation between MH and LH was .87 ($\sigma_r = .03$).

In view of these high correlations the data for only one of the response measures, total time on target, will be presented. As may be seen in Fig. 1, the five groups were very nearly equal initially and did not differentiate from one another even after 45 trials.

The .2-sec. delay group appeared slightly superior to the others, especially during the No Tone period, but this advantage was not significant.

An extended trend analysis (**2**) was run for each of the three response measures and the two conditions of information feedback. In each case there were no significant differences between the groups. Only the between-individual and the over-all linear and quadratic terms were significant.

The significant over-all linear and quadratic terms indicate a nonlinear increase in performance with practice. The significant between-individual means, slopes, and quadratic terms reflect individual differences. The absence of any significant between-groups terms indicates the ineffectiveness of the delayed feedback variable.

Discussion

There appears to be but one conclusion to draw from the present experiment. Delay of information feedback is not an effective variable in the acquisition of skill in rotary pursuit tracking. Apparently the visual feedback is sufficient to account for *S*'s performance and the effect of any additional feedback is negligible. An alternative hypothesis does emerge, however, in view of at least two differences between the present study and that by Reynolds and Adams. Both differences probably would result in higher motivation in the *S*s of the latter study. Their *S*s (*a*) served in groups which probably fostered competition, and (*b*) were Air Force basic trainees. It is quite probable that many of the trainees suspected that their performance could determine, to some degree at least, their assignment to an Air Force career field. Since career-field assignment is a topic of considerable discussion and importance to a basic trainee, it seems likely that the click was highly rewarding as well as informative. At least it seems probable that it was more rewarding than the tone was to the college students of the present study. Accordingly, the alternative hypothesis is that the effectiveness of delay of information feedback is a function of the level of motivation of the *S*s.

Summary

Five groups of 15 men each learned a rotary pursuit task for 45 trials. During the first 30 trials, *S*s heard a 1000-cps tone after being on target continuously for one of five durations. These delay of information feedback intervals were .0, .2, .4, .8, and 1.6 sec. During the last 15 trials no tone was heard.

The results were negative in that different delays of information feedback did not significantly affect performance. These results were unlike those of Reynolds and Adams (**3**). An explanation in terms of different motivational effects of information feedback was offered.

References

1. Archer, E. J., Kent, G. W., & Mote, F. A. Effect of long-term practice and time-on-target information feedback on a complex tracking task. *J. exp. Psychol.*, 1956, **51**, 103–112.
2. Grant, D. A. Analysis of variance tests in the analysis and comparison of curves. *Psychol. Bull.*, 1956, **53**, 141–154
3. Reynolds, B., & Adams, J. A. Motor performance as a function of click reinforcement. *J. exp. Psychol.*, 1953, **45**, 315–320.

(Received November 18, 1957)

Journal of Experimental Psychology
Vol. 56, No. 4, 1958

A SCALE OF APPARENT INTENSITY OF ELECTRIC SHOCK [1]

S. S. STEVENS, A. S. CARTON, AND G. M. SHICKMAN

Harvard University

The purpose was to construct a ratio scale for the subjective magnitude of the sensory response that results when 60-cycle current is passed through the fingers, and to compare this scale with the category scale obtained under similar circumstances. Although the sensation of electric shock is interesting in its own right, a main point of this study was to determine whether Ss can make quantitative judgments (method of magnitude estimation [5]) on a continuum with which they have had no experience. This question is relevant to the argument sometimes made that magnitude estimations are dependent on the prior association of sensed magnitude with known stimulus conditions. Another point of interest concerns the quantitative relation between sensed magnitude and stimulus magnitude. Is this relation similar in form to the one that obtains on prothetic continua, i.e., is it a power function? And if so, what is the exponent?

A further question concerns the relation of this function to the loudness function observed in patients whose auditory nerve was stimulated directly by an electric current. The fact that these two functions of electrical stimulation turn out to be rather similar, but quite different from the loudness function determined with acoustic energy, may have a bearing on the role played by the sense organ in determining the rate of growth of sensation magnitude.

[1] Supported by a grant from the National Science Foundation and by a contract with the Office of Naval Research (Project NR142-201, Report PNR-209). Reproduction for any purpose of the U. S. Government is permitted.

RATIO-SCALE STUDIES

Experiment I

Method.—Three separate experiments were run to determine the ratio scale for apparent intensity of electric shock. The first (conducted by Carton) was in the nature of a preliminary exploration. The apparatus consisted simply of a Variac voltage control whose output was stepped up by a "power transformer." A 100,000-ohm resistor in series with the output of this transformer made the system behave approximately as a constant-current source, with the result that variations in the resistance of the fingers had negligible effect on the current. The current was measured with the aid of a vacuum-tube voltmeter by noting the voltage drop across a 10,000-ohm resistor in place of the fingers. Two fingers (usually the first and third) of one hand were immersed up to the first joint in salt solutions in two glass jars, on the inside of which were fastened metal electrodes. Fingers with cuts or abrasions were not used. The *E* administered the shock by closing a mercury switch for a period of about 1 sec.

Written instructions were given as follows: "I am going to present you with a series of currents of different intensities. Your task is to tell me how intense they feel by assigning numbers to them. The first intensity we will call '10.' Your task will be to estimate the intensity of each stimulus in the series. Try to assign numbers proportional to your subjective impression of each stimulus. Try not to worry about being consistent, but try to give the appropriate number to each stimulus regardless of what you have called some previous stimulus."

The *E* then presented the standard stimulus (.79 milliamp.) and reiterated that its apparent intensity should be called "10." He then presented the other stimuli twice each in an irregular order—a different order for each of 10 Ss (graduate students and staff members). The "standard" value of the stimulus was also presented as a stimulus to be judged along with the other values.

Results.—In Fig. 1 are shown the medians (filled triangles) and the geometric means (unfilled triangles) of the magnitude estimations. It is clear from these results that the magnitude estimations approximate a straight line in a log-log plot and, therefore, that subjective intensity is approximately proportional to the value of the current raised to a power. The exponent is about 3.5.

Actually the triangles in Fig. 1 suggest that at the high current values the growth of the sensation becomes less steep. This results mainly from the fact that 4 of the 10 *S*s gave the same numerical estimate for the top two stimuli. As a check on this matter, these two stimuli were presented one after the other to each of the four *S*s and they were asked whether the two currents felt the same. The answer was No in every case. It appears, therefore, that the decreased slope in Fig. 1 may not be representative, although it is not improbable that the curve turns over at high levels.

It should be noted that the stronger currents used in this experiment were very disagreeable. People do not like to be shocked, even by weak stimuli, and they find strong stimuli decidedly aversive. Nevertheless, they can apparently judge subjective intensity with fair consistency.

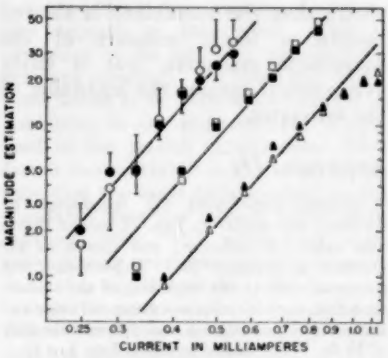

FIG. 1. Three experiments on the estimation of the apparent intensity of electric current. The filled points are medians; the unfilled points are geometric means. The vertical lines through the circles mark off the interquartile ranges.

Experiment II

Method.—For Exp. II (conducted by Shickman) the apparatus was modified to provide more exact control of the duration of the shock (.5 sec) and to relieve *S* of the tension-generating process of having to wait for the onset of the current without knowing when to expect it. A key was provided by which he could initiate its onset whenever he was ready. The standard stimulus, to be called "10," was set at .37 milliamp. and this stimulus was presented at the beginnings of both the first and second runs. Otherwise the procedure was the same as in Exp. I. It should also be noted that the series resistor was increased to 250,000 ohms in this and Exp. III and IV. The *S*s were 15 graduate students and staff members, two of whom were females. Four *S*s had served in Exp. I. Previous participation seems to have no systematic effect on the form of the resulting magnitude functions in these experiments.

Results.—The medians (filled circles) and the geometric means (unfilled circles) are shown in Fig. 1, where it may be noted that the slope of the function is a little steeper than that determined in Exp. I. The weakest current used in Exp. II was so close to threshold that on almost one-third of the presentations *S* reported no sensation. These blank responses were omitted in computing medians and geometric means. Some impression of the variability among the judgments in experiments of this sort is conveyed by the vertical lines through the circles in Fig. 1. These lines mark the interquartile ranges. The variability is perhaps a little greater than it is in comparable experiments on loudness (cf. **5**). The general agreement between the medians and the geometric means in Fig. 1

shows that the variability is skewed against a linear measure of the magnitude estimates, but is fairly symmetrical against the logarithm of the estimates.

Experiment III

Method.—Experiment III (conducted by Stevens) was similar to Exp. II, except that a new value (.5 milliamp.) was chosen as the standard to be called "10." This standard was presented only at the beginning of the session. As before, each stimulus was presented twice and the order of presentation was different for each of 10 *Ss*. Four *Ss* had served in Exp. I or II.

Results.—The medians (filled squares) and the geometric means (unfilled squares) in Fig. 1 confirm the approximate power function found in the two earlier experiments. One point is rather far off the line, however. This fact points up one of the problems that arise in magnitude estimation, namely, the "round-number" tendency. The weakest stimulus used was sufficiently close to what the median *S* would normally call "1" so that half the *Ss* reported this convenient value for at least one of their judgments. On the other hand, the circles in Fig. 1 show that over this low end of the stimulus range the slope is not as steep as the squares would suggest. The use of different standards, as in these three experiments, often provides a convenient means of detecting distortions due to preferences for particular numbers.

Note, incidentally, that two of the filled circles fall opposite the same value on the ordinate. The value "5" is a popular estimate when the standard is called "10" and the stimulus seems roughly half as strong as the standard.

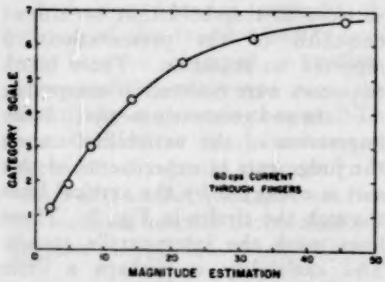

FIG. 2. Category scale of apparent intensity of shock plotted against the scale determined by magnitude estimation. The curvature is typical of category scales on prothetic continua.

CATEGORY-SCALE STUDY

Experiment IV

Method.—If the perceptual continuum of electric shock is a prothetic continuum, the category scale should be nonlinearly related to the ratio scale obtained by magnitude estimation (10). In order to test this hypothesis, a category scale was erected with the same stimuli as those represented by the squares in Fig. 1. The *S* was first presented with the weakest current (.33 milliamp.) and told to call it "1." He was then given the strongest stimulus (.84 milliamp.) and told to call it "7." He was then asked to rate the various stimuli on a scale from 1 to 7. In other words, the procedure was that of the classical experiments in which people try to assign successive numbers to values that seem equidistant from one another, i.e., to use the numbers to mark off equal-appearing intervals. Each of 8 *Ss* judged each stimulus twice in a different irregular order. Six of the *Ss* had served in one or more of the experiments on magnitude estimation.

Results.—In Fig. 2 the average category assignments are plotted against the scale of magnitude estimation, i.e., against the scale determined by the middle line in Fig. 1. As seems to be universally true on prothetic continua, the category scale is concave downward when plotted against the scale of subjective magnitude.

DISCUSSION

Although the direction of the curvature of the category scale seems to be invariant over all prothetic continua, the

degree of the curvature can be altered by the spacing of the stimuli. Evidence from numerous other experiments (10) suggests that the curvature in Fig. 2 could be decreased by spacing the stimuli closer together over the high end of the range, or increased by spacing them closer together over the low end of the range. On the other hand, the form of the ratio scale of subjective magnitude, obtained by magnitude estimation, is essentially invariant with stimulus spacing (3). This fact suggests that, for the measurement of subjective magnitude, the ratio scale is superior to the inconstant and demonstrably nonlinear category scale. Actually, these category scales turn out not to be proper interval scales of subjective magnitude, even though they are obtained by methods that ask *S* to try to partition a continuum into equal-appearing segments. As explained elsewhere (10), the asymmetry of *S*'s responsiveness to differences—the fact that a given difference near one end of the scale is more impressive than the same difference near the other end of the scale—makes it impossible for him to partition a prothetic continuum without a systematic bias. It is mainly this bias that produces the curvature in Fig. 2.

The foregoing facts and argument may help to explain why Swartz's experiment (11) on electric shock applied to the tooth to produce a sensation of pain resulted in a scale that was approximately linear with current. He used a method of bisection and, except for a large hysteresis effect (cf. 7), his results are not unlike those obtained by category scaling. If plotted against current, the points in Fig. 2 would fall reasonably close to a straight line. This fact, however, should not be taken as evidence that subjective intensity is proportional to current. It may be that, if Swartz had asked his *S*s to make direct estimations of the apparent magnitude of the pain they experienced, the results would have produced an accelerated function like that in Fig. 4.

Adaptation.—When the current is set at a fairly high value and left on for a few seconds a certain amount of adaptation appears to take place. This is evidenced by the fact that the threshold becomes elevated. On the other hand, there seems to be little if any change in sensitivity to the short-duration shocks used in the present experiments. This fact is demonstrated in Fig. 1 where we note that the same currents used as the standards were not judged significantly different from "10" when they were subsequently presented for judgment. It is also significant that there was no systematic tendency for the judgments to be lower on the second run than on the first run through the series.

A subjective scale.—By an adjustment of the vertical scales of the three sets of data in Fig. 1, the points can be brought into approximate coincidence and plotted as a single line in a log-log plot. This has been done in Fig. 3. With both the abscissa and the ordinate adjusted to the same modulus, the function turns out to be very steep. The over-all slope is close to 3.5, which indicates that the apparent intensity of the shock grows approximately as the 3.5 power of the current.

At this point it seems in order to define a unit for this subjective scale and give it a name. The unit may conveniently be defined as the subjective impression experienced by the typical *S* when a 60-cycle current of .3 milliamp. is passed through his fingers. Since this unit is a kind of "subjective ampere," it will be called a *samp*. In terms of this unit, the apparent intensity *S* of electric shock is approximated by the formula

$$S = 67.6\, I^{3.5}$$

where *I* is current in milliamperes.

A plot of this function in linear coordinates produces the sharply accelerating curve shown in Fig. 4. Although the evidence of the foregoing experiments makes it clear that the growth of sensation becomes more and more rapid as the current is increased, the exact formula should be regarded more as an approximation than as a precise rule. The exponent is probably

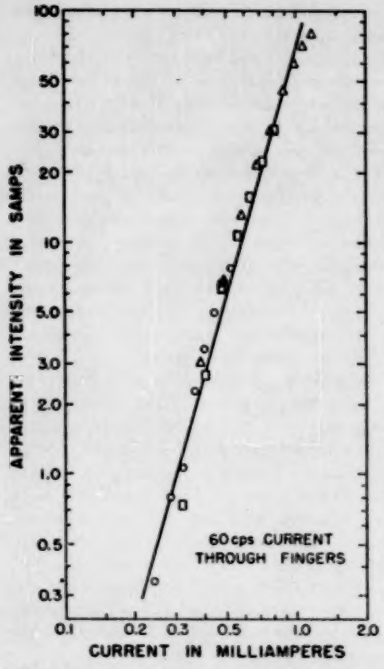

FIG. 3. The geometric means from Fig. 1 were adjusted vertically to fall along a common line in a log-log plot. The ordinate value of 1 samp is made to coincide with a current of .3 milliamp. The slope of the line is 3.5.

correct in order of magnitude, but it has not yet been checked by other ratio-scaling methods (9). It is also evident that the function may in some respects depart appreciably from a power function. There is a tendency, possibly significant, for the function to be slightly concave downward in a log-log plot. Over the low current levels the exponent appears to be larger (of the order of 4 or more) than it is when the current is stronger.

Nevertheless, the function for shock is certainly different from that for some of the other sensory continua such as loudness (7), which grows as the .3 power of the sound energy, or the .6 power of the sound pressure. This difference suggests a rather basic difference in the physiological mechanisms involved. With loudness there is a process of "compression" in the sense that the psychological magnitude grows less rapidly than the stimulus magnitude. With shock there is a process of "expansion": the psychological magnitude grows more rapidly than the stimulus magnitude.

Electrical stimulation of the auditory nerve.—Since the apparent intensity produced by a current through the fingers grows so rapidly with stimulus magnitude, it becomes a matter of interest to inquire whether a similar rapid growth may occur when other nerves are stimulated electrically.

In a study of the electrophonic effect (2), ears lacking tympanic membranes were stimulated by means of an electrode placed in the middle-ear cavity. Eleven patients heard pure tones, and seven heard only a buzzing noise whose character was more or less independent of the frequency of the stimulating current. It is these latter cases that are of interest here. Since it is clear that other nerves, such as the facial and the vestibular, were stimulated in some of these patients, it is reasonable to conclude that the auditory nerve itself was sometimes directly affected by the cur-

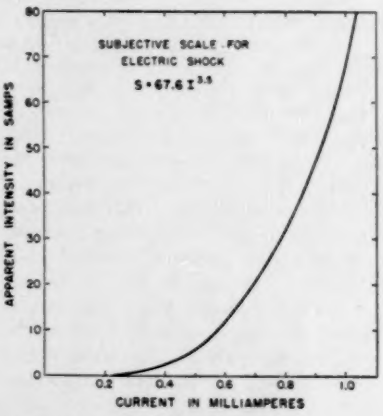

FIG. 4. The function in Fig. 3 is plotted in linear coordinates.

rent. An unpatterned stimulation of the auditory fibers would account for the fact that a noise rather than a tone was heard.

Fortunately for the present concern a brief study was made of the dependence of loudness on current in some of the ears whose auditory nerves seem to have been stimulated electrically. An acoustical stimulus was applied to the opposite (normal) ear, and the patient adjusted its loudness to match that of the noise produced by the current in the operated ear. This procedure demonstrated that the loudness of the noise grows very rapidly as the current is increased. For example, when the decibel increase in the acoustic stimulus required to keep the loudness equal in the two ears is plotted against the decibel increase in the current applied to the nerve, the slope of the line is almost exactly the same as that in Fig. 3.

This suggestive similarity is not the whole story, however, for the relation of interest is that between current through the nerve and loudness in subjective units (sones). The sone function (**4, 6**) can be used to convert the sound pressure levels generated by the earphone into values proportional to loudness in sones, and these values can be plotted against the voltage applied to the operated ear. Some illustrative values are shown by the circles in Fig. 5. In this log-log plot the loudness is seen to grow very rapidly. The function is actually about four times as steep as the sone function determined acoustically, but it does not seem to be quite as steep as the line in Fig. 3.

On the other hand, as Jones, Stevens, and Lurie (**2**) emphasized, the data in Fig. 5 underestimate the rate of growth of loudness under nerve stimulation. The reason is that, during the time required to make the loudness match, the patients reported that the loudness of the noise fell off quite noticeably, especially for the stronger currents. Without this rapid "adaptation" the measured growth of loudness would presumably have been steeper.

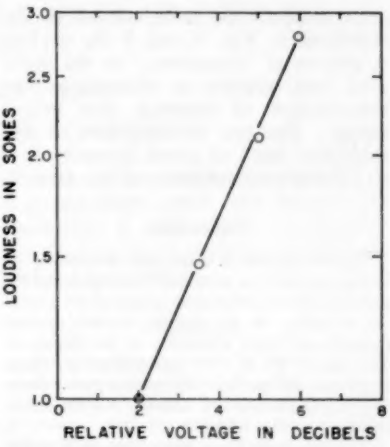

FIG. 5. Showing approximately how loudness grew as a function of the strength of an electrical stimulus applied to the auditory nerve in a patient whose eardrum had been removed. The current was delivered by an electrode placed in the middle ear. This function is about four times as steep as that determined with an acoustical stimulus applied to a normal ear.

If this fact is allowed for, it would appear that both the sensation of loudness and the sensation of shock in the fingers grow in approximately the same manner as the current is increased. The implication is that the "compression" observed when a method such as magnitude estimation is used to determine the sone scale with acoustic energy is a function of the end organ, not of some higher center in the nervous system. When the electrical stimulus by-passes the end organ and acts directly on the nerve, there seems to result the same process of "expansion" noted in the case of shock (cf. **8**).

Although it is hazardous to draw conclusions about sensory processes in human beings from observations made on nerve preparations in lower animal forms, it is not without interest that the rate of firing of an axone may also grow rapidly with the strength of an exciting current. In at least some instances the frequency of response in a single fiber seems to

grow in about the same manner as the functions in Fig. 3 and 5 (1, p. 76). A process of "expansion," in the sense used here, appears to characterize the transduction of electrical into neural energy. Perhaps the steepness of the subjective scale of shock intensity is a straightforward reflection of this fact.

Summary

By the method of magnitude estimation, *Ss* having no previous experience in judging electric shock made numerical estimations of the apparent intensity of an electric current applied through salt-water electrodes to the fingers of one hand. To a first approximation, these magnitude estimations determine a power function in which subjective intensity S is related to current I by $S = 67.6\ I^{3.5}$.

The exponent 3.5 is larger than any of the exponents previously determined for some 15 other prothetic continua, but it is of the same order of magnitude as that observed when the auditory nerve is stimulated electrically. The difference between the exponents for acoustical and electrical stimulation suggests that the "compression" observed under acoustical stimulation is a function of the sense organ, not of the central nervous system.

A category scale was determined by having *Ss* judge various currents on a scale from 1 to 7. Like the usual category scale on a prothetic continuum, this scale is concave downward when plotted against the ratio scale determined by magnitude estimation.

References

1. Brink, F. Excitation and conduction in the neuron. In S. S. Stevens (Ed.), *Handbook of experimental psychology*. New York: Wiley, 1951. Pp. 50–93.
2. Jones, R. C., Stevens, S. S., & Lurie, M. H. Three mechanisms of hearing by electrical stimulation. *J. acoust. Soc. Amer.*, 1940, **12**, 281–290.
3. Stevens, J. C. Stimulus spacing and the judgment of loudness. *J. exp. Psychol.*, 1958, **56**, 246–250.
4. Stevens, S. S. The measurement of loudness. *J. acoust. Soc. Amer.*, 1955, **27**, 815–829.
5. Stevens, S. S. The direct estimation of sensory magnitudes—loudness. *Amer. J. Psychol.*, 1956, **69**, 1–25.
6. Stevens, S. S. Concerning the form of the loudness function. *J. acoust. Soc. Amer.*, 1957, **29**, 603–606.
7. Stevens, S. S. On the psychophysical law. *Psychol. Rev.*, 1957, **64**, 153–181.
8. Stevens, S. S. Measurement and man. *Science*, 1958, **127**, 383–389.
9. Stevens, S. S. Problems and methods of psychophysics. *Psychol. Bull.*, 1958, **55**, 177–196.
10. Stevens, S. S., & Galanter, E. H. Ratio scales and category scales for a dozen perceptual continua. *J. exp. Psychol.*, 1957, **54**, 377–411.
11. Swartz, P. A new method for scaling pain. *J. exp. Psychol.*, 1953, **45**, 288–293.

(Received November 27, 1957)

Journal of Experimental Psychology
Vol. 56, No. 4, 1958

RELIABILITY OF JUDGMENTS OF FIGURAL COMPLEXITY [1]

LOIS LAWRENCE ELLIOTT

Air Force Personnel and Training Research Center

In a recent investigation of judgments of figural complexity of random shapes, Attneave (**2**) related scaled complexity values to measurable physical characteristics of the stimuli. Arnoult included the complexity condition of judgment in an experiment which also obtained ratings of random shapes in terms of size, familiarity, and esthetic qualities.[2] Both experiments were administered to populations of basic trainee airmen at a military installation and employed similar experimental procedures. Those shapes common to both stimulus populations had highly correlated complexity scale values. Other than this check there has been no investigation of the reliability of these judgments.

The present experiment was designed to study the generality and reliability of judgments of figural complexity in terms of responses from diverse subject populations and different methods of judgment and scaling. It also tests the assumption that results obtained from a sample of random shapes may be generalized to other samples from the stimulus domain (**3**). Although planned as a group experiment, the design also permitted a preliminary analysis of consistency of individual judgments of figural complexity.

METHOD

Subjects.—One hundred *S*s served in each of two experiments (total $N = 200$). The *S*s were student officer cadets, recent college graduates, receiving training prior to being assigned to pilot school. They differed from the basic trainee *S*s of Arnoult's and Attneave's work in forming a more homogeneous group with regard to educational experiences, aptitudes, and physical qualifications. A group of 10 or 15 was tested at one time.

Apparatus.—The study was conducted in a large room equipped with dark shades. The stimuli were projected onto a 5×7-ft. screen by means of 2×2-in. slides. The shapes varied in size but subsumed, on the average, approximately 3° of visual angle. All stimuli were members of the 72-item nonsense shape population employed by Arnoult. One half of the shapes were constructed by Method 1 as outlined by Attneave and Arnoult (**3**), while the other half was constructed by their Method 2. Examples of the stimulus materials are shown in Fig. 1. In Part A, 20 shapes were selected to cover the range of complexity values obtained by Arnoult. All possible pairs (190) of the stimuli were presented once each. These same 20 stimuli plus an additional 20 shapes chosen at random from the Arnoult population

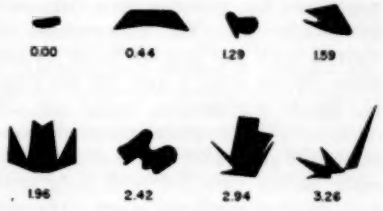

FIG. 1. Samples of the 72-item nonsense shape population with complexity scale values obtained by Arnoult.

[1] This research was done under ARDC Project No. 7740, Task No. 27106, in support of the research and development program of the Air Force Personnel and Training Research Center, Lackland Air Force Base, Texas. Permission is granted for reproduction, translation, publication, use and disposal in whole and in part by or for the United States Government.

[2] M. D. Arnoult. Prediction of perceptual responses from structural characteristics of the stimulus. (In preparation.) The writer wishes to thank Dr. Arnoult for making available his stimulus materials and the scaled complexity values obtained in his experiment.

formed the stimulus materials for Part B. Experiment II replicated both parts of Exp. I. None of the stimuli used in Part A of Exp. I (paired comparisons) appeared in the corresponding part of Exp. II. In Part B (rating scale), however, the two experiments had eight stimuli in common.

Procedure.—In Part A, 20 shapes were judged in complexity by the paired-comparisons method. In accordance with earlier studies, the term "complexity" was not defined for *S*. Instead, five practice pairs of shapes were presented; the first pair was chosen to have maximum difference in complexity (as defined by the Attneave equation [2]), the following pair had less difference in complexity, etc. In almost every instance *S* asked no further questions. The 190 text pairs were presented each with a 5-sec. exposure time. One order of pair presentation was used for half of the *S*s in each experiment, the other half viewing the stimulus pairs in the reverse order. The *S*s recorded their responses on IBM answer sheets. The entire procedure occupied about 28 min. After a 5-min. break, Part B was administered. The *S*s rated 40 shapes presented individually (5-sec. exposure) on a 5-point rating scale (the same procedure employed by Attneave [2] and Arnoult[3]). Response categories were "very complex," "complex," "medium," "simple," and "very simple."

Results and Discussion

Scaled complexity values were computed for Part A by Thurstone's Case V (6) method and for Part B by Attneave's Graded Dichotomies technique (1). The spread of scale values by both methods was zero to approximately 4.0. Reliability was assessed by examining the correlation between the two different measures of the same stimulus materials. The procedure has recently been discussed by Noble (8). The correlations of Part A with the Arnoult scale values were .94 for Exp. I and .98 for Exp. II (each correlation based on 20 items). These correlations demonstrate the generality of the complexity scale across two different judgment and scaling methods when different subject populations are tested. Correlations of Part B (40 shapes) with the Arnoult scale values were .95 for Exp. I and .96 for Exp. II, indicating the reliability of the rating scale values for different subject populations. When Parts A and B were intercorrelated (20 shapes), the coefficients were .92 and .98 for the separate experiments and .92 for the pooled data (40 shapes). Thus, reliability for different judgment and scaling methods is demonstrated for the same subject population. And finally, all three sets of correlations indicate reliability between subsamples of the total shape population.

The result substantiates an assumption underlying current efforts to develop a psychophysics of form perception. Attneave and Arnoult (3) have assumed that one may sample from any population of shapes, or stimulus domain, which is characterized by determinate statistical parameters and may generalize results based on the sample to the population. In Part A of both experiments, a sample of only 20 shapes produced results which were consistent with those obtained from the group of 72 shapes. The same was true of samples of 40 shapes rated in Part B.

Investigation of reliability of the complexity scale on a group basis was the primary purpose of this experiment. In handling the data, it became apparent that useful information could be extracted concerning consistency of individual judgments. One would prefer for each *S* a measure of the width of the complexity scale interval which could be discriminated with stated probability, i.e., an interval analogous to a difference limen. In the literature there appears to be no precedent for approaching this problem, although Gerard and Shapiro (4) have discussed a method for determining degree of inconsistency in paired-comparisons data. Therefore, two different descriptive techniques were examined. The first was based on paired-comparisons data, the second on rating scale responses. Each had certain drawbacks as well as advantages.

Method 1.—The term "consistency" was defined in the following manner: all pairs of shapes having one member of the pair in common were

[3] See footnote 2.

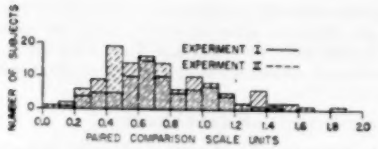

FIG. 2. Distribution of mean intervals of inconsistency for paired-comparisons judgments.

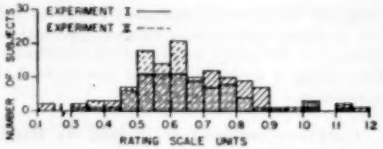

FIG. 3. Distribution of variances computed by Method II (rating-scale judgments).

arrayed so that the comparison shapes fell in rank order from high to low complexity (scale values). An *S* would display perfect consistency of judgment for those pairs if all comparison shapes of rank value higher than *n* were judged as more complex than the standard (shape common to all pairs) and all shapes with rank value lower than *n* were judged as less complex. For any standard, the interval of inconsistency was defined as

$$\left[\frac{M_r + M_{r+1}}{2} - \frac{L_{r-1} + L_r}{2}\right]$$

where:

L_r was the scale value of the comparison shape with highest rank to which *S*'s response was "less complex than standard" and such that every comparison shape of higher rank was judged "more complex."

L_{r-1} was the scale value of the shape with rank immediately higher than L_r, to which *S*'s response was "more complex."

M_r was the scale value of the shape with lowest rank value to which *S* responded "more complex than standard" and such that every shape of lower rank was judged less complex.

M_{r-1} was the scale value of the shape with rank value immediately lower than M. The *S*'s response was "less complex." For the paired-comparisons method it is possible to compute as many such intervals as there are stimuli. In this study, however, it seemed feasible to compute 10 such intervals, using as standard the shapes with scale values ranking 6 through 15. A mean interval of inconsistency was computed for each *S*. Distribution of the mean interval of inconsistency for both experiments is shown in Fig. 2; medians are .65 and .70 complexity scale units, or approximately one sixth the total range of the complexity scale.

In evaluating this degree of consistency one must recognize that the over-all complexity scale range was limited by the nature of the particular stimulus domain. With specification of other parameters, different and probably larger ranges of complexity scales could be obtained. For the present situation, *S*s appear to have demonstrated a relatively high degree of consistency of judgments. The advantage of Method 1 is the fact that the computed interval is directly related to the complexity scale continuum. On the other hand, it suffers from lack of independence of the comparison arrays for the different "standards." Further, it equates the importance of all possible "L_r" shapes regardless of the response to the L_{r+1}, L_{r+2}, etc., shapes. It could be considerably improved by filling in both halves of the square matrix by experimental design, obtaining multiple observations per cell, and clarifying its mathematical properties. Although the author is not aware that this exact approach has been previously discussed, it falls in line with work of Mosier (**7**) on "individual variability in test theory" and Glaser (**5**) on inconsistency of response.

Method 2.—For each *S* a variance estimate was computed based on the difference between *S*'s rating for each stimulus item and the group mean rating for that item (40 stimulus items).[4] An error not larger than 1% may have been introduced into the results because the group item mean included the judgment of the particular *S* under consideration. Results are graphed in Fig. 3.

The variance measure computed by this method is not related to the complexity scale continuum. It indicates degree, but not direction, of consistent agreement with mean group ratings (not scale value) of shapes considered individually. Compared to Method 1, its mathematical properties are clearer and more in accord with traditional descriptive techniques.

Correlations of Methods 1 and 2 by *S*, for both experiments, are not significantly different from zero (−.18 and −.14). This should not be surprising since Method 2 measures the extent of *S*'s agreement with the group judgment of *isolated shapes* while Method 1 considers judgment of a pair of shapes in relation to all other pairs. As experimental techniques, Method 2 is appropriate for evaluating con-

[4] Analysis of independent data (**2**, Footnote 2) has indicated that scale values obtained by the Attneave (**1**) method are approximately linearly related to mean ratings.

formance with group judgments while Method 1 (in a more sophisticated version, perhaps) is demanded to study sensitivity of discrimination along a multidimensional psychophysical continuum.

Certain questions are raised by these results. Could *S*s be trained to improve their consistency of judging figural characteristics such as complexity? Are there consistent individual differences in the relative weighting of specific physical characteristics—i.e., do some people "pay more attention" to symmetry than to number of sides in judging complexity? How would changing the properties of the stimulus domain affect consistency of judgment? Is ability to discriminate complexity of random shapes associated with ability to recognize and identify complex shapes? These problems await further research.

Summary

Two groups of 100 *S*s each judged the complexity of different samples of random shapes by both the paired-comparison and rating-scale methods. Intercorrelations and correlations of these scales with the Arnoult scale values all fell above .92. Judgments of figural complexity were shown to be consistent for different populations using the same scaling method, for the same subject population using different methods, and between different shapes from the same population. Two methods for evaluating individual consistency of judgment were described.

References

1. Attneave, F. A method of graded dichotomies for the scaling of judgments. *Psychol. Rev.*, 1949, **56,** 334–340.
2. Attneave, F. Physical determinants of the judged complexity of shapes. *J. exp. Psychol.*, 1957, **53,** 221–227.
3. Attneave, F., & Arnoult, M. D. The quantitative study of shape and pattern perception. *Psychol. Bull.*, 1956, **53,** 452–471.
4. Gerard, H. B., & Shapiro, H. N. Determining the degree of inconsistency in a set of paired comparisons. *Psychometrika*, 1958, **23,** 33–46.
5. Glaser, R. A methodological analysis of the inconsistency of response to test items. *Educ. psychol. Measmt*, 1949, **9,** 727–739.
6. Guilford, J. P. *Psychometric methods*. New York: McGraw-Hill, 1954.
7. Mosier, C. Psychophysics and mental test theory: II. The constant process. *Psychol. Rev.*, 1941, **48,** 235–249.
8. Noble, C. E. The meaning-familiarity relationships. *Psychol. Rev.*, 1953, **60,** 89–98.

(Received December 13, 1957)

Journal of Experimental Psychology
Vol. 56, No. 4, 1958

GENERALIZATION OF AN INSTRUMENTAL RESPONSE WITH VARIATION IN TWO ATTRIBUTES OF THE CS [1]

SHELDON H. WHITE [2]

Iowa Child Welfare Research Station

The stimuli involved in a discrimination usually differ on one physical dimension but have common values on others. If the difference between the distinguishing components of the stimuli is made larger, the discrimination is learned more easily. This can be explained by the principle of stimulus generalization, as used in discrimination learning theories (**3, 6**).

To date, studies of generalization have always involved the gradual alteration of one component of a CS while holding the others constant. Theory, however, has recognized the possibility that more than one component may be so varied and, by postulating summative rules for generalized reaction tendencies, can predict that discrimination learning will be superior with multiple, as opposed to single, dimensions of variation between discriminanda. This has been confirmed in several studies (**1, 2, 8**).

This experiment was an attempt to demonstrate that novel stimuli differing from a training stimulus along two physical dimensions will elicit fewer generalized responses than will novel stimuli which differ only on one of those dimensions.

[1] A portion of a dissertation submitted to the State University of Iowa in partial fulfillment of the requirements for the Ph.D. degree. The writer wishes to express his gratitude to Charles C. Spiker, who directed this research, and to Don Lewis, for valuable suggestions. John Peterson gave considerable aid in construction and maintenance of the apparatus.

[2] Now at the University of Chicago.

METHOD

Apparatus.—The apparatus was a modification of one developed by Spiker (**7**) to study generalization with children. It allows *S* to respond by repeated pulls of a handle during a trial.

The stimuli were presented by an experimental model of the Hunter Cardmaster, a device for the timed presentation of cards. This is a gray metal box, 14 × 14 × 14 in., with a stimulus aperture 6 in. wide and 3 in. high. Within, a conveyor system slides a 6¼ × 3½-in. card from the bottom of a stack and presses it against the aperture. Black cloth curtains part to either side to reveal the card to *S*. When the curtains close, the card is returned to the top of the stack and a new one brought up. Each stimulus exposure lasted 3.8 sec., an interval which preliminary testing indicated would allow a sufficiently high rate of response for gradations in generalization to manifest themselves. Each intertrial interval lasted 3.5 sec.

The opening of the curtains closed a switch in the circuits of an impulse counter and a marble delivery device; these circuits were then completed when *S* pulled the response handle ¾ in. towards himself. (A 3-lb. pull would close the microswitch.) Thus, when the curtains were open, each pull would register on the impulse counter and cause a marble to be delivered. When they closed, the counting and marble delivery stopped. A hidden button enabled *E* to lock out marble delivery during any given trial.

The handle protruded from the lower right-hand corner of a black wooden box on which the Cardmaster rested, and varied from waist to chest height for different *S*s. Tubing, housed in a black box to the left, held a reserve of about 300 marbles. These were delivered into a plastic pail at *S*'s left.

Stimuli.—The stimuli were 3 × 5-in. sheets of Munsell colored paper, covered with library tape and mounted on white plastic cards. In Table 1, the colors are schematized in the Munsell notation (**5**), which is of the form: Hue Value/Chroma. The colors were illuminated by a 100-w. bulb about 2 in. above, shielded to prevent glare. The blinds of the various schoolrooms used were drawn, to make this the principal source of illumination of the stimuli.

Procedure and Ss.—The children were told

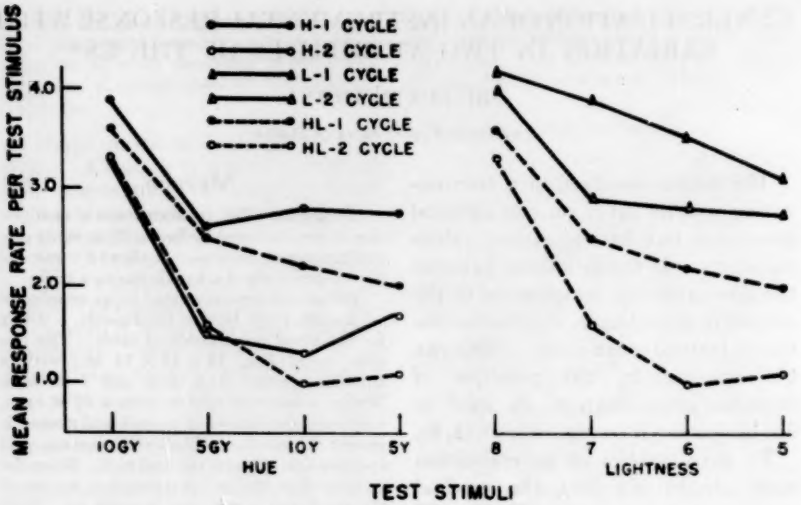

FIG. 1. Mean response rates per test stimulus in Groups H, L, and HL, as a function of variation in hue and lightness.

they were to play a game, and that the way to win it was to fill a pail with marbles. The *E* pulled the handle to two exposures of the training stimulus, and *S* was encouraged while pulling to six more. Occasionally, when *S* had not given three or more responses on each of two consecutive trials, he was given up to four more. Marbles were delivered on all pretraining trials.

All *S*s were then given 16 reinforced trials with the training stimulus and, without pause, a sequence of 8 nonreinforced test trials distributed among 12 reinforced presentations of the training stimulus. Either 1, 2 ,or 3 reinforcements separated one test trial from the next.

The transition to testing produced a change in incidental stimulation. The color used in training, being always the same, was presented by opening and closing the curtains on one card. During the last training trial, *E* closed a switch causing the Cardmaster thereafter to change cards at the end of each trial. Card changing produced a sound noticeable to *S*. The first trial of the testing sequence was, therefore, a reinforced presentation of the training stimulus, in an attempt to counteract any change in behavior which might be produced by the new sound.

There were three testing groups. Group H had four test stimuli: the training color and three others differing in hue, but of the same lightness. Group L had test stimuli varying in lightness, but of the same hue. The novel stimuli for Group HL differed from the training color in both hue and lightness. Each HL color contained a Group H hue and a Group L lightness (Table 1).

The *S*s were kindergarten children in the Cedar Rapids public schools[3] who met a criterion of not more than one error on Plates I–VI of the

TABLE 1

MUNSELL VALUES OF COLORS USED AS STIMULI VALUE (LIGHTNESS)

Hue	8/	7/	6/	5/
10GY	*10GY 8/6*	10GY 7/6	10GY 6/6	10GY 5/6
5GY	5GY 8/6	5GY 7/6		
10Y	10Y 8/6		10Y 6/6	
5Y	5Y 8/6			5Y 5/6

Note.—Group L's test stimuli along row, HL's along diagonal, H's in column. The training stimulus was 10GY 8/6.

[3] The writer wishes to thank Hale C. Reid, Director of Curriculum and Instruction for permission to use as *S*s children of the public schools of Cedar Rapids, Iowa. Ruth Phillips, principal of the Grant Wood School, and Florence Gritzner, principal of the Lincoln School, provided facilities for, and made pleasant, the running of *S*s.

TABLE 2

ANALYSES OF VARIANCE COMPARING GROUPS H VS. HL, AND L VS. HL, ON NUMBER OF RESPONSES TO THEIR FOUR TEST STIMULI

Source	*df*	H vs. HL		L vs. HL	
		MS	*F*	*MS*	*F*
Between *Ss*	47				
Groups (G)	1	9.69	<1	150.00	9.12**
Error (b)	46	11.04		16.44	
Within *Ss*	336				
Stimuli (S)	3	64.34	30.21***	47.04	23.76***
Cycles (C)	1	83.44	71.93***	45.37	18.52***
G × S	3	2.40	1.13	3.81	1.92
G × C	1	1.63	1.41	1.26	<1
S × C	3	3.11	2.80*	2.57	2.40
G × S × C	3	0.03	<1	0.64	<1
Error (w)	322				
$Error_1$ (w)	138	2.13		1.98	
$Error_2$ (w)	46	1.16		2.45	
$Error_3$ (w)	138	1.11		1.07	
Total	383				

* $P < .05$.
** $P < .01$.
*** $P < .001$.

ISCC Pseudoisochromatic Colorblindness Test, administered before an experimental session. The data for 24 children were not used. Fourteen *Ss*—four from H, four from L, and six from HL—were interrupted by mechanical failure of the apparatus. Three HL *Ss* were dropped because of procedural errors, and three other HL *Ss* because an extremely rapid rate of response overfilled their marble container. One *S* from H, and another from L, were eliminated because of distractions within the school. Two children failed to meet the color-blindness criterion. There remained a total of 72 *Ss*, 24 in each group.

RESULTS

During training, Groups H, L, and HL averaged, respectively, 6.4, 6.5, and 6.6 pulls per trial. Analysis of variance did not reveal a significant difference between the groups.

The left half of Fig. 1 shows the mean number of pulls per test stimulus, cycles given separately, for Groups H and HL. There is little difference between them on their first two test values, but Group H gives more responses than Group HL on the two furthest points. For both groups, responses to the 10GY (training) value in test trials were markedly greater than to the three novel test stimuli, which did not differ greatly in numbers of responses elicited. There is a decrease in response rate from the first to the second cycle of test trials. These effects, except for the difference between groups, were found to be significant in an analysis of variance (4) (Table 2).

An analysis of variance with the factors shown in Table 2, but using only the data of the three novel test stimuli, was performed. In this analysis, only the cycles effect was significant, at the .001 level.

From the right half of Fig. 1, it is seen that (*a*) Group L gave more generalized responses to all four of its test stimuli than did Group HL to its corresponding stimuli; (*b*) a gradient of generalized responses sloping away from the training value is indicated;

(*c*) there is a decrease in responses from the first to the second cycle. Analysis of variance indicated these effects to be significant (Table 2).

A repetition of this analysis using only the data of the three novel test stimuli of each group again showed these three effects, and no others, to be significant.

Discussion

The fact that more *S*s were eliminated from Group HL than from the H and L groups together raises the possibility that some sampling bias was inadvertently introduced. The most serious contamination would be in the selection of slower responders into Group HL, since evidence of this and other (**8**) studies tends to show low positive correlations between rates of response during acquisition and test trials. However, the acquisition data shows that all three groups gave about the same rate of response during training, and suggests that the groups' differences in generalization were probably not produced by selection of *S*s.

The hypothesis of the present study can be predicted by considering each attribute of the training stimulus to be a separately conditionable component, with its habit strength summating with that of other attributes to produce the habit strength for the compound CS. Hull's laws of primary stimulus generalization and habit summation (**3**) would then predict that: (*a*) the three groups should respond identically on test trials to the training color; (*b*) the HL generalization gradient should be below those of the other two groups; and (*c*) the HL gradient should be steepest.

For Prediction *a*, we may note that the H, L, and HL groups averaged, respectively, 3.6, 4.1, and 3.5 responses over their two test trials to the training color, with nonsignificant differences. For Prediction *b*, we note that Group L was significantly above, and Group H tended to be above, Group HL in numbers of responses on their generalization test trials. With regard to Prediction *c*, it is evident in Fig. 1 that the differences between Groups H or L and HL are greater at the furthest test point than at the training point, as would be expected. However, the Groups × Stimuli effect fails to approach significance in either analysis of Table 2.

The above discussion of theory and test has disregarded the extinction effects which resulted from a counterbalanced testing procedure, and which affect the generalization curves. The comparison of test stimuli within *S*s was based, as it usually is, upon a consideration of the great economy of *S*s which the method provided. In terms of theory, at least, the predicted effects should usually manifest themselves despite distortions produced by this method. The prediction which is most vulnerable to distorting effects of extinction is *c*, that of differing slope for the HL, as compared to the H or L gradients.

Summary

Three groups of 24 kindergarten children were given 16 trials of training to pull a response handle freely during 3.8-sec. presentations of a colored stimulus. Two cycles of generalization test trials were then interspersed among reinforced presentations of the training stimulus. Group H was tested on the training stimulus and three novel stimuli of the same lightness, but having graded differences in hue. The novel stimuli of Group L were like the training stimulus in hue but differed in lightness, while those of Group HL differed in both hue and lightness.

The three groups did not differ in their rates of response during training. Group HL gave fewer generalized responses to its test stimuli than did Groups H or L to theirs, but only the comparison between L and HL was significant.

References

1. Eninger, M. U. Habit summation in a selective learning problem. *J. comp. physiol. Psychol.*, 1952, **45**, 604–608.
2. Eriksen, C. W., & Hake, H. W. Multidimensional stimulus differences and accuracy of discrimination. *J. exp. Psychol.*, 1955, **50**, 153–160.
3. Hull, C. L. *A behavior system.* New Haven: Yale Univer. Press, 1952.

4. LINDQUIST, E. F. *Design and analysis of experiments in psychology and education* Boston: Houghton Mifflin, 1953.
5. MUNSELL, A. H. *A color notation.* (8th ed., edited and rearranged.) Baltimore: Munsell Color Co., 1936.
6. SPENCE, K. W. The differential response in animals to stimuli varying within a single dimension. *Psychol. Rev.*, 1937, **44**, 430–444.
7. SPIKER, C. C. The stimulus generalization gradient as a function of the intensity of stimulus lights. *Child Develpm.*, 1956, **27**, 85–98.
8. TAYLOR, J. E. The effects of differential positive-negative cue differences on the changing strength of response tendency. Unpublished doctor's dissertation, State Univer. of Iowa, 1953.

(Received December 2, 1957)

Journal of Experimental Psychology
Vol. 56, No. 4, 1958

FACILITATION OF MIRROR-IMAGE WORD IDENTIFICATION BY MIRROR-IMAGE PERCEPTUAL SET

SIDNEY S. CULBERT

University of Washington

No thorough investigation has been made of the perceived similarity between a mirror image and its obverse, though the matter has received some incidental attention. Attneave (**1**) in his insightful study on dimensions of similarity admitted (p. 525) that he had overlooked the factor of symmetry in his first experiment and had to introduce it into his system as a new dimension when *S*s unexpectedly reported one parallelogram as similar to another which was approximately the mirror image of the first. Henle (**3**) found the structural equivalence of a letter and its mirror reversal to provide a useful device in her investigation of visual form perception. Chou (**2**) pointed out that mirror reversal of a passage (an example of what he called "orientation of the first order") made for less difficulty in reading than when individual letters were reversed ("orientation of the second order"). The only explanation he offers is that "the word-Gestalt has been changed" in the latter case (**2**, p. 55). The present study is aimed at investigating the effect which was noted in these reports. More specifically the question may be put: Is the mirror image of a figure with left-right orientation (such as a word or numerical expression) more easily identified if it appears on a field which is perceived as a mirror image? Situations of this type occur occasionally in everyday life. We see words or sentences reflected in mirrors; we see a reversal of the name of the soda fountain or beer tavern as we assuage our thirst and glance idly at the plate glass front. Does the fact that the mirror image is properly oriented in accordance with our expectations make it easier to read than the same mirror image would be in a situation in which a mirror image would not fit our perceptual set?

Three experiments were carried out to investigate this question. The first two were designed to measure the recognizability of a mirror-reversal word[1] when seen on a reversal of a previously presented field as compared to the word's recognizability when seen on the nonreversed field. In Exp. I the reversal of the field was abrupt and arbitrary. In Exp. II the reversal occurred in a "reasonable" fashion, the field was moved horizontally through a 180° arc in order to give *S*s a "back" (i.e., reversed) view of the field. In Exp. III, the stimulus material consisted of anagrams composed of reversed letters. A comparison was made of the time necessary for their solution when viewed directly on a screen as compared to the time required when their mirror images on a screen were viewed in a mirror.

EXPERIMENT I

Method.—The reversal of the field and the sudden appearance of the reversed word on the reversed field is most easily arranged by single-frame photography, the completed film pro-

[1] The term "word" will be used to refer to both words and numerical expressions used. See Table 1.

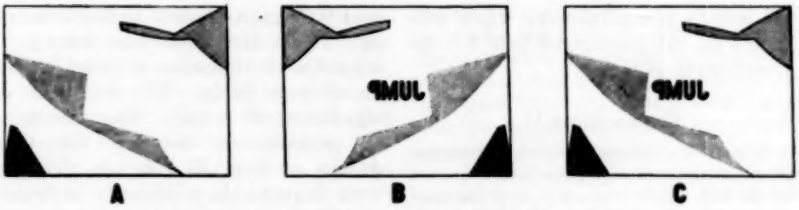

FIG. 1. The three fields used in Exp. I and II: (A) Obverse field without stimulus word, (B) reversed field with stimulus word, (C) obverse field with stimulus word.

viding the stimulus material actually presented to *S*.

The field was a group of nonsense figures made of various brightnesses of gray paper glued onto a transparent rectangle of plate glass 28 cm. long by 21 cm. high. The area in the central part of the rectangle was left free of figures (Fig. 1, A). The glass plate was held in a rectangular frame, one end of which was mounted on a vertical rod. The field was presented to *S* for 8.25 sec. (132 frames at 16 frames/sec.), after which the stimulus situation differed for experimental *S*s and control *S*s.

In the case of experimental *S*s the field instantly reversed (to the mirror image of itself) and there simultaneously appeared in the occupied central portion a mirror reversal of a four-letter word which remained visible for .75 sec. (Fig. 1, B). (The length of the interval was based on preliminary experiments which indicated that about .5 to 1 sec. represented the range of recognition thresholds for the reversed words under these experimental conditions.) Both field and word then disappeared from the screen and were replaced by four words in obverse, one of which was the obverse of the reversed (stimulus) word which had been presented on the field and the others the obverse of words similar to the stimulus word (Table 1). The *S* was requested to guess which one of the four had appeared on the field.

In the case of control *S*s the procedure was identical except that the field was not reversed during the presentation of the reversed word (Fig. 1, C).

Each of 17 experimental *S*s and 17 control *S*s was presented with 12 different reversed words, each word being followed by a four-choice group (in obverse) containing the correct word. The number of correct answers represented *S*'s score. Scores would be expected to fall between 3 (chance) and 12 (all correct).

Results.—There was no significant difference between the groups, the mean score of the experimental group being 9.47 and that of the control group 9.35.

It may be concluded that under the conditions of the experiment an instantaneous substitution of a reversed field for the previously presented field does not facilitate the recognizability of a reversed word appearing on the field simultaneously with its reversal as compared to

TABLE 1

GROUPS OF WORDS PRESENTED TO ALL *S*s IN EXP. I AND II

(Asterisks, not shown to *S*s, indicate correct work or numerical expression in each group)

PLUM	ETCH	FIFE	SINK	BOND*	GROW*
JUMP*	STUB	FARE	SIGH	SNOB	WORN
PAUL	CUTE*	ERIE	KISS	BORN	ROSE
JUNE	TRUE	FIRE*	SICK*	KNOB	GOWN
5809	6721*	82100	TANK	7859	PAY$6
9U85	6/27	0.128	TINT*	8597	$6055*
5890	1276	$2155*	INTO	7958*	00068
9085*	1270	$1255	KNIT	T958	$6000

the word's recognizability when presented on the unchanged field for the same length of time.

Experiment II

Method.—Experiment II required a somewhat more elaborate photographic procedure than did the first. Instead of a mere substitution of a reversed field for its obverse, the situation was arranged so that *S* would see the field swing through a semicircle. The right side of the frame holding the glass plate was hinged to a vertical rod. On the first view, *S* saw the field at the right angles to his line of regard and to the left of the rod. The field remained motionless for 6 sec. It then swung toward *S* and continued through a 180° arc which brought the frame into a position again at right angles to *S*'s line of regard but now to the right of the vertical supporting rod. The movement from left to right was not constant, but slow at first and accelerating rapidly until the extreme right position was reached, when the picture disappeared. The appearance of acceleration was accomplished by the use of a specially constructed timing apparatus which synchronized the position of the field with the flashing of a Strobolux and the movement of film in a 16-mm. camera. The purpose of the acceleration was to create an illusion of smooth movement from the starting position by showing the field in 36 positions during the first 90° of its swing (until the field was edge-on for *S*) but permitting only a brief glimpse of the field during the last half of the swing by photographing only 12 positions past the edge-on position. The first half of the swing took 2.25 sec., the second half .75 sec. The situations for both experimental and control *S*s were identical until the edge-on position was reached—both saw the obverse field motionless for 6 sec. and then swinging to the edge-on position for 2.25 sec. During the second half of the swing the experimental *S*s were presented with a reversal of the field, i.e., the orientation one would expect when the field had passed the edge-on position. Control *S*s were presented with the obverse—the same field which had appeared during the 8.25 sec. before the edge-on position was reached. In both cases the reversed word appeared in the unoccupied central area of the field during the .75 sec. after the edge-on position was passed. As in Exp. I, *S* was asked to indicate which of four words had appeared during the .75-sec. period.

Results.—The mean score of the 17 *S*s who were presented with the reversed words on the reversed fields was 9.59 as compared to mean score of 7.53 for the 17 *S*s who were presented with the same reversed words on obverse fields. The t of 3.29 is significant ($P < .01$). The difference in performance between the two groups of Exp. II indicates that *S*s were better able to identify reversed words on mirror reversals of previously presented fields than the same reversed words on the obverse of previously presented fields.

Experiment III

Method.—The experimental material consisted of the mirror images of the following anagrams: KDCE, LEJL, PEREC, NEECS, ZEREEF, EKEN, SNDBE, GLEED, SPEJE, EFRERP. These particular (block) letters were used because all others are bilaterally symmetrical, the mirror image being identical with the obverse (Q is not available since it always appears with U). Lower-case letters were avoided since the mirror images of some are identical with the obverse of others. The anagrams (mirror images) were presented consecutively by means of a transparency projector to individual *S*s. The time taken by *S* to solve each anagram was recorded by an electric timer controlled by *E*. The projected image was 10 cm. high and appeared at an optical distance of 4 m. Each *S* was given a sample anagram to be sure he understood the instructions which were simply to solve each anagram as presented and give the solution aloud.

Experimental *S*s sat beside the screen, looking into a mirror (44 cm. long by 30 cm. high) placed at a distance of 2 m. from *S* (and from the screen). The anagrams were projected onto the screen in obverse and consequently were seen as reversed (i.e., mirror images) by the experimental *S*s. Control *S*s were seated at a distance of 4 m. from the screen and slightly to one side in order to establish the same angle of line of regard as occurred in the case of experimental *S*s. The anagrams were projected directly as mirror images for control *S*s. The stimulus conditions were similar for the two groups except for location of the *S*s and the experimental variable of a mirror in the case of the experimental *S*s. There were 40 *S*s in each group.

Results.—A pretest of the experimental procedure on 20 *S*s indicated that an occasional *S* could be ex-

pected to fail to solve one or possibly more of the anagrams in the 6 min. maximal time allowed on each. For this reason the score taken to represent the performance of each *S* was his median time for the solution of the 10 anagrams. The mean of the medians of the experimental *S*s (mirror group) was 18.67 sec.; that of the control *S*s, 28.27 sec. The difference between the means of the two groups was significant ($t = 2.71$, $P < .01$). Since no *S* failed to solve more than three anagrams, the mean of the middle four values was also available for each *S*. Using this measure of each *S*'s performance as the basis for calculation of group means leads to a *P* value similar to that based on the medians.

Among the experimental *S*s, there were 12 failures to solve anagrams, involving 11 *S*s. In the control group there were 28 failures scattered among 19 *S*s. The difference between number of individuals in the two groups who failed to solve all anagrams is not significant ($P > .05$), but is in the direction that would be expected from the comparison of group means. An inspection of the "fast" ends of the distributions shows no obvious difference between the groups.

Discussion

In both Exp. I and Exp. II all *S*s were presented with the obverse field *without* the reversed word for 8.25 sec. followed by the field (obverse or reverse) *with* the reversed word for .75 sec. In Exp. II, however, the field was swinging during the last 3 sec. of the total (9 sec.) period. During the last .75 sec. the swinging field occupied 12 different positions to the right of the edge-on position, the first few making such small angles with the line of regard that the reversed word was not legible even when stationary in those positions. This means that all *S*s in Exp. II had considerably less than .75 sec. to identify the reversed word. Even in the last few positions, when the angles with the line of regard were large, the reversed word was continually changing its retinally projected length, no two presentations being the same. Nevertheless, *S*s presented with reversed fields during this period did as well as *S*s in either group of Exp. I; the reversed words on reversed fields swinging through a 90° angle from an edge-on to a right angle position were read as well as the same reversed words on stationary reversed fields and as well as reversed words on stationary obverse fields, the time (.75 sec.) being the same in all three situations. It might be argued that the facilitating effect of a reversed field found in Exp. II did not occur in Exp. I because *S*s in Exp. I presented with the reversed field did not recognize it as the obverse of the previously presented field. This possibility was not formally checked on the actual *S*s of Exp. I, but of 20 practice *S*s, asked to talk about the situation after the first trial, all reported such recognition.

Expectation, in the sense of being able to say "the word to be identified is always backwards," has no explanatory value for the results of Exp. II, since all *S*s saw the words to be identified in no other orientation. But the swinging field seems to have built up a "perceptual set" that facilitated the identification of the reversed word only if the expected orientation, a mirror image of the field, actually developed.

It might reasonably be contended that the inferior performance of the *S*s presented with the obverse fields could be explained by assuming that the unexpected orientation after the edge-on position was passed served to distract them from the task. In Exp. I, however, the sudden reversal did not seem to distract the experimental *S*s—at least their performance was not poorer than that of the control group. It is, of course, possible that the reversed field did distract *S*s but that a facilitating effect of the reversed field canceled the decrement in performance. A further

bit of evidence is available through comparison of performance of *Ss* on the first six words with their performance on the last six. Though the difference between control and experimental *Ss* might be expected to decrease with habituation to the distraction, no such decrease is found.

The results of Exp. III indicate that an effect similar to that appearing in Exp. II may be found not only in recognition but in a situation where mirror-image letters must be in some way "manipulated"—whether in visual imagery, through verbal mediation, or by means of some other process. Experiment III also indicates that the "perceptual set" (whatever it is) does not depend on reversal of a previously presented field but may be brought about through use of a device, a mirror, which *Ss* already know to be a "field reverser."

Summary

Three experiments were performed to determine whether a reversed or mirror-image field would facilitate the identification of reversed words or the "manipulation" of reversed letters. Experiment I showed no such facilitation when the reversed field abruptly replaced its obverse at the time of appearance of the reversed word, this situation being compared to the continuation of the obverse field during the presentation of the reversed word. Experiment II showed that such facilitation does occur if the originally presented field is swung through a 180° arc, becoming a mirror reversal (the expected orientation) during the last 90° of its swing. In Exp. III it was found that anagrams composed of block letters could be solved more easily when viewed in a mirror than when their mirror images were seen directly on a screen.

References

1. Attneave, F. Dimensions of similarity. *Amer. J. Psychol.*, 1950, **63**, 516–556.
2. Chou, S. K. Gestalt in reading Chinese characters. *Psychol. Rev.*, 1930, **37**, 54–70.
3. Henle, M. An experimental investigation of past experience as a determinant of visual form perception. *J. exp. Psychol.*, 1942, **30**, 1–22.

(Received December 16, 1957)

Journal of Experimental Psychology
Vol. 56, No. 4, 1958

THE EFFECT OF A STIMULANT AND A DEPRESSANT DRUG ON A MEASURE OF REACTIVE INHIBITION [1]

S. N. SINHA,[2] CYRIL M. FRANKS,[3] AND P. L. BROADHURST

Institute of Psychiatry (Maudsley Hospital), University of London

Eysenck's drug postulate (**6**) suggests that depressant drugs increase cortical inhibition and that stimulant drugs decrease it. It is the purpose of the present experiment to test this prediction as directly as possible by obtaining a gradient of reactive inhibition (I_R) in the rat, and then endeavoring to alter the slope of that gradient by the use of drugs.

Alternation behavior in a simple T maze (**9**) was chosen as the measure of I_R, and was manipulated by inserting a variable number (from 0 to 3) of previous turns in the maze pattern before the final choice point. Specifically, the prediction is that the amount of alternation—that is, turning in the direction opposite to that of the previous turns—will be a function of the number of those turns (**8, 10, 12**) and, further, that this function will itself be increased by a depressant drug and decreased by a stimulant drug. The availability of *Ss* from the Maudsley Reactive and Nonreactive strains, selectively bred for emotional defecation (**4**)—the high defecators displaying a higher drive level in fear-producing situations than the low defecators (**1, 2**)—allows a subsidiary prediction. Following Kimble (**11**), it would be predicted that the emotionally reactive *Ss* should show less alternation than the nonreactive ones, due to their greater tolerance of I_R.

METHOD

Apparatus.—The apparatus basically consisted of a simple wooden T maze with arms 3½ in. wide and 4 in. high, roofed with hardware cloth and painted flat gray. Each arm was 1 ft. long externally. The whole was enclosed within a rectangular runway of identical interior dimensions and construction, but which measured 22.5 in. by 34.5 in. externally. As shown in Fig. 1, this rectangular runway communicated with the T maze only at the lower end of the upright of the T. The alley forming the other side of the rectangular runway, opposite this junction, was permanently divided into two halves. Thus, by suitable manipulation of the series of temporary blocks within this outer runway, *Ss* could be run to a food reward either from the foot of the T (Starting Position 0), or from either of the lower corners of the outer rectangle, left or right (Starting Position 1), or from either of the upper corners (Starting Position 2), or from either side of the permanent block on the upper side of the rectangle (Starting Position 3). An *S* would thus have made 0, 1, 2, or 3 turns, respectively, before entering the final T, and these turns could be either to the right or to the left. The additional distance

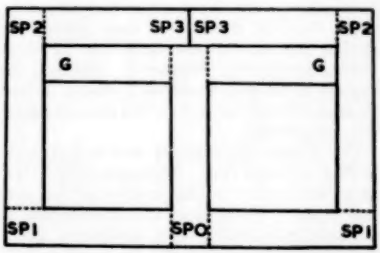

FIG. 1. A floor plan of the maze used, showing the permanent block in the outer runway (solid line) and the movable blocks (dashed lines) used in conjunction with the different starting positions (SP) and goal arms (G).

[1] We are indebted to H. J. Eysenck for many helpful suggestions, and to Riker Laboratories, Ltd., Loughborough, Leics., England, for supplying the Meratran used.

[2] Now at Patna University, Bihar, India.

[3] Now at Neuro-Psychiatric Institute, Princeton, New Jersey.

run in each case was not exactly equal, however, because of the nature of the design.

Two 60-w. fluorescent tubes above the apparatus gave illumination within it of 20 to 22 ft.-c. at floor level. It was enclosed within a dome of cheesecloth in order to eliminate extra-maze clues.

Subjects.—A total of 32 rats, 16 males and 16 females, were selected from the fifth generation of the selectively bred, emotionally reactive strain, and another 32 from the nonreactive strain. They had been reared under uniform conditions and given a standardized version of Hall's open-field test of emotionality at 100 days of age, as described elsewhere (**3**). Only *Ss* scoring three or more fecal boluses per day on this test were assigned to the Reactive group, and only *Ss* not defecating at all to the Nonreactive group. They averaged 226.1 ($SE \pm .62$) days of age when the food deprivation for the purposes of the present experiment was begun. It was established gradually, only reaching the final value of 1 hr. feeding per day after 10 days, and 5–8 days before testing began. The *Ss* were therefore 20–23 hr. hungry at the time of testing throughout the experiment.

The 64 *Ss* were randomly assigned, within the sex and emotionality dichotomies, to the four starting positions. There were thus four replications of the $2 \times 2 \times 4$ factorial design.

Procedure.—On the two days before maze running began, *Ss* were given preliminary adaptation training in the apparatus by allowing them to feed on wet mash for 3 min. in each of the two goal boxes at either end of the arms of the T. The *Ss* assigned to Starting Position 3 (Group 3) were started first and given 10 trials from Starting Position 0 on the first day, 10 from Starting Position 1 on the next, 10 from Starting Position 2 on the next, and finally 30 trials from Starting Position 3 on the following three days. The Starting Position 2 *Ss* (Group 2) were begun one day later so that they completed their 30 trials from their appropriate starting position on the same day as those assigned to Starting Position 3. Similarly, *Ss* assigned to Starting Positions 1 (Group 1) and 0 (Group 0) were started on the two successive days, respectively.

The *Ss* were permitted to feed on wet mash for 10 sec. on each trial. Throughout this training *Ss* were forced to one or other of the goal arms in the T by blocking the opposite side at the junction according to a Gellerman series. They were, apart from the *Ss* starting from Position 0, also started from the appropriate left or right starting position according to another Gellerman series. Thus *Ss* were trained to run equally, not only to both of the goal arms of the simple T in the center, but also from both the right or the left starting position. Times to the nearest .1 sec. were recorded by stop watch, and analysis by *t* test of these data from the 30 training trials for *Ss* from Groups 1 through 3 shows no significant difference in mean running times from either right or left starting position to either right or left goal arm. Group 0, however, ran significantly faster to the left goal arm, but, as will be seen later, this does not appear to be reflected in a heightened preference for this side. In addition, in no case did the mean times from the last training day (10 trials) show a significant difference from those during the subsequent part of the experiment. It may, therefore, be concluded that these training procedures, designed to establish equal asymptotic habit strengths (sH_R) in running to either right or left goal from the various starting positions, were, in general, successful.

All *Ss* were then given the same alternation test lasting two days, which consisted in allowing *S* a free choice in the T, both goal arms now being left open. This was done on eight trials on two successive days, and each such free-choice trial was followed by a forced trial with one or other of the arms blocked. Thus, a total of eight trials, four free alternating with four forced, was given each day with a constant intertrial interval, despite the differences in starting position. The starting position side was selected as before—for *Ss* other than those in Group 0—by use of a modified Gellerman series, but the goal side which was blocked on the forced trials was in each case determined by *S*'s own choice on the previous free-choice trial, so that each *S* was forced away from the side previously chosen. This was a further precaution against the development of position habits.

This two-day alternation test was repeated twice under different drug conditions—that is, each *S* was tested under each of the three conditions, stimulant or depressant drug and placebo. The six possible orders of these three conditions were distributed over the 16 *Ss* in each group as widely as possible, so that each of the four subgroups therein (e.g. male reactive) is represented at least twice and not more than four times, and that over all *Ss* each order was assigned to at least 10 *Ss*.

The depressant drug used was amobarbital sodium (Sodium Amytal) which has been successfully used in other tests of Eysenck's drug postulate (e.g., **7**). The dose was 15 mg./Kg. of body weight. For the stimulant drug the amphetamine group usually employed was eschewed particularly because of its anorexic side effects, and pipradrol hydrochloride (Meratran) was selected. This is a relatively new central

stimulant, but without anorexic effects (**5**). The dose used was 10 mg./Kg. The placebo was 10 cc./Kg. of distilled water, and the drugs mentioned were also given in this volume of distilled water. All injections were made intraperitoneally with a No. 26 record needle, 38.3 (±.26) min. before testing began.

Results

During each free-choice trial an alternation response was scored only when *S* turned to the goal arm of the T in the direction opposite to that of its previous turns in the maze on that same trial. An *S* with a position habit would thus not vitiate the result because it would, by always turning to one side and being *started* from each side four times, score only four out of the possible eight alternations—that is, chance level. Alternation responses for *Ss* from Starting Position 0, without previous turns, were scored in a comparable way, the justification for which is the assumption that I_R may persist from one trial to the next in this simple situation (**14**). In this case, an alternation response was defined as a turn on free-choice trial to the goal arm of the T in the direction opposite to that to which it was forced on the previous trial. For the first (free-choice) trial on any day, the scoring convention adopted was to regard the last (forced) trial on the previous day as influential. The possibility of artifacts in the data caused by position habits among these *Ss* is more serious as it is not possible to reduce their effect to chance level in the manner described above. Moreover, by our scoring, an *S* which always turned to the same side would receive the maximum alternation score of eight, since it would in each case be forced to the opposite side on the intermediate trials. Examination of the data, however, showed that no *S* in this group responded consistently in this way.

The results of a four-way analysis of variance of the alternation scores, after Pearson and Hartley's test (**13**) had disclosed no significant inhomogeneity of variance, are presented in Table 1. It will be seen that only the variance estimates for the main effects and interactions not involving Drugs were calculated by using the *Ss*' error term; and those for all other sources by using the *Ss* × Drug interaction. This procedure is consequent upon the experimental design used in which each *S* was tested under each drug condition. The *F* ratios for the Starting Position and Drug main effects are shown in this way to be significant, but note must be taken of the significant Starting Position × Drug interaction. The nature of this interaction may be

TABLE 1

Analysis of Variance of Alternation Scores and Running Times

Source	*df*	Alternation Scores		Log Times	
		MS	*F*	*MS*	*F*
Starting position (P)	3	40.0	9.7***	1.73	46.8***
Drugs (D)†	2	37.0	16.0***	0.08	8.0***
Emotional reactivity (E)	1	0.1	0	0.01	0.2
Sex (S)	1	8.3	2.0	0.73	20.0***
P × D†	6	8.4	3.6**	0.02	2.3*
P × E	3	1.7	0.4	0.01	0.3
P × S	3	4.8	1.2	0.14	3.7*
D × E†	2	0.5	0.2	0.01	1.1
D × S†	2	1.6	0.7	0	0
E × S	1	0	0	0	0
P × D × E†	6	0.8	0.4	0.02	1.7
P × D × S†	6	0.9	0.4	0	0
P × E × S	3	0.6	0.1	0.05	1.5
D × E × S†	2	0.4	0.2	0.01	0.5
P × D × E × S†	6	1.5	0.6	0	0
Residual					
Ss × D	96	2.3		0.01	
Ss	48	4.1		0.04	
Total	191				

* $P = .05$.
** $P = .01$.
*** $P = .001$.
† *Ss* × D residual used for *F*.

TABLE 2

BREAKDOWN OF ALTERNATION SCORES

Drug Condition	Starting Position								All	
	0		1		2		3			
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Amobarbital (Amytal)	3.7	2.4	5.6	0.9	6.7	1.5	6.9	0.8	5.7	2.0
Placebo	3.7	1.7	4.7	1.4	6.3	1.5	5.6	1.5	5.1	1.8
Pipradrol (Meratran)	4.3	2.5	3.7	0.9	4.8	1.4	4.1	1.7	4.2	1.7
All	3.9	2.2	4.7	1.3	5.9	1.7	5.5	1.8	5.0	1.9

seen in Table 2 and Fig. 2. Each increase in the number of previous turns yields an increase in the mean alternation score, which is significant ($P < .05$ or beyond by 1-tail t test) in each case except that between Starting Positions 2 and 3 where in fact a decrease occurs. However, the differences in the mean alternation scores for the different drugs are both significant in the predicted direction beyond the .01 level by t test. A separate analysis showed that the difference between the mean scores for Ss when started from left or from right starting position does not reach significance.

The results of an analysis of variance of the time scores derived from the eight free-choice trials, after the log transformation made had reduced the highly significant inhomogeneity of variance to around the .05 level, are of less interest, serving principally to confirm the successful manipulation of two of the independent variables in the experiment. Thus, the significant main effect for Starting Position reflects the successive increases in time taken to run from the four positions (the mean log times being .26 for Starting Position 0, .49 for Position 1, .58 for Position 2, and .71 for Position 3, with the differences between the successive means significant at the .02 level or beyond by 1-tail t test) and that for Drugs reflects the increases in running times under the three drugs in the expected order pipradrol, placebo, and amobarbital (mean log times of .47, .51 and .54 respectively; the successive differences being significant at the .05 level or beyond by 1-tail t test). But a significant sex difference also appears—females run faster than males. The difference between their respective mean log times of .49 and .57 is significant beyond the .001 level by 2-tail t test. In addition, the over-all picture is complicated by weakly significant interactions between Starting Position on the one hand, and Sex and Drugs on the other. These interactions may be summarized by saying that the sex difference becomes progressively more pronounced with an increase in the number of previous turns, but the drug differences become less pronounced. A further separate analysis showed that the differences between the mean times for free and forced trials during the alternation tests does not reach significance.

It will be noted that there is no support whatsoever for the prediction of differences in I_R with respect to the Emotional Reactivity variable. The comparable lack of significant effects in the analysis of the running times confirms that strain differences in drive intensity did not operate in this situation.

DISCUSSION

It may be concluded from these results that the prediction made regarding the effect on alternation behavior of a stimulant and a depressant drug in the doses used are confirmed. The amount of alternation is increased by the depressant, amobarbital, and decreased by the stimulant, pipradrol, with the placebo group having an intermediate value. Thus, another ex-

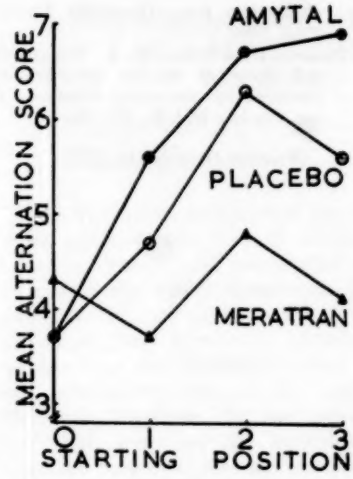

FIG. 2. The mean alternation score (maximum possible = 8) for each starting position group of 16 *Ss* under the three drug conditions.

ample, of a somewhat different kind and using very different *Ss*, is added to the growing support (6) for the essential soundness of Eysenck's drug postulate.

Whether or not it is accepted that these findings also support Eysenck's identification of I_R as being the important element in this postulate depends on the view of alternation behavior adopted. The fact that it has been possible to demonstrate a gradient of alternation in this experiment lends further support to the view that alternation behavior is a valid measure of I_R. The failure of alternation to increase with an increase in the number of previous turns from two to three parallels work with lower organisms and may similarly be regarded as a dissipation of I_R during the increased time taken to run from the farthest starting position (8).

SUMMARY

In a test of Eysenck's drug postulate using alternation behavior in the rat as a measure of reactive inhibition (I_R), it was predicted that a stimulant drug, pipradrol, would decrease alternation and a depressant, amobarbital, would increase it. In a $2 \times 2 \times 4$ design, 64 *Ss* of the two sexes and from the two Maudsley Reactive and Nonreactive strains were assigned to four different starting positions in a maze designed to produce a gradient in the number of alternation responses shown. They were then tested under the influence of the two drugs and a placebo.

The results show that the predictions relating to drug action are confirmed at a satisfactory level of significance.

REFERENCES

1. BROADHURST, P. L. Emotionality and the Yerkes-Dodson Law. *J. exp. Psychol.*, 1957, **54**, 345–352.
2. BROADHURST, P. L. A "Crespi effect" in the analysis of emotionality as a drive in rats. *Brit. J. Psychol.*, 1958, **49**, 56–58.
3. BROADHURST, P. L. Determinants of emotionality in the rat: II. Antecedent factors. *Brit. J. Psychol.*, 1958, **49**, 12–20.
4. BROADHURST, P. L. Studies in psychogenetics: The quantitative inheritance of behaviour in rats investigated by selective and cross-breeding. *Bull. Brit. psychol. Soc.*, 1958, **34**, 2A. (Abstract)
5. BROWN, B. B., & WERNER, H. W. Pharmacologic studies on a new central stimulant, (2-Piperidyl) benzhydrol hydrochloride. *J. Pharmacol.*, 1954, **110**, 180–187.
6. EYSENCK, H. J. *The dynamics of anxiety and hysteria: An application of modern learning theory to psychiatry.* London: Routledge & Kegan Paul, 1957.
7. FRANKS, C. M., & LAVERTY, G. Sodium amytal and eyelid conditioning. *J. ment. Sci.*, 1955, **101**, 654–663.
8. GROSSLIGHT, J. H., & TICKNER, W. Variability and reactive inhibition in the mealworm as a function of determined turning sequences. *J. comp. physiol. Psychol.*, 1953, **46**, 35–38.
9. HOUSE, B. J. The relationship of place and turn alternation to reactive inhibition. *J. genet. Psychol.*, 1956, **89**, 183–197.
10. IWATA, K., & WATANBE, M. Alternate turning response of *Armadillidium vulgare*: 3. Effect of preceding turning. *Annu. anim. Psychol., Tokyo*, 1957, **7**, 56–60.
11. KIMBLE, G. A. Evidence for the role of motivation in determining the amount

of reminiscence in pursuit-rotor learning. *J. exp. Psychol.*, 1950, **40**, 248–253.

12. Lepley, W., & Rice, G. Behavior variability in paramecia as a function of guided act sequences. *J. comp. physiol. Psychol.*, 1952, **45**, 283–286.
13. Pearson, E. S., & Hartley, H. O. *Biometrika tables for statisticians*. Vol. 1. Cambridge, Eng., Cambridge Univer. Press, 1956.
14. Zeaman, D., & House, B. J. The growth and decay of reactive inhibition as measured by alternation behavior. *J. exp. Psychol.*, 1951, **41**, 177–186.

(Received December 16, 1957)

Journal of Experimental Psychology
Vol. 56, No. 4, 1958

THE MAGNITUDE OF BINOCULAR SUMMATION AS A FUNCTION OF THE METHOD OF STIMULUS PRESENTATION [1]

GEORGE COLLIER AND PHILIP KUBZANSKY

University of Missouri *Boston City Hospital, Harvard Medical School*

The binocular summation experiment provides one way of deciding between central and peripheral accounts of the visual threshold. Occurrence of a significantly lower binocular than monocular threshold implies that the threshold events are not exclusively retinal in origin. However, Pirrenne (**8**) has pointed out that, even on the assumption that the two eyes operate independently, the binocular threshold should be lower than the monocular by the rule for combination of independent but nonmutually exclusive probabilities. Thus a demonstration of summation requires that monocular-binocular threshold differences exceed the value expected on the combination of independent probabilities.

In previous studies all possible outcomes have been obtained: no differences between monocular and binocular thresholds, the differences expected by the combination of independent probabilities, and differences exceeding that expected by the combination of independent probabilities (**4**). This diversity of results led to a study (**4**) which carefully controlled the variables most often related to the failure or success in demonstrating summation, e.g., fixation, differential accommodation, differential pupil size, and stimulation of noncorresponding retinal points. In order to evaluate Pirenne's hypothesis directly and to test the possible peripheral origin of the nonindependence of successive responses observed under some circumstances in threshold measurements, a single brightness method of stimulus presentation was used. A brightness, determined for each *S* from initial method of limits data, was used for that *S* in all conditions. This yielded a P_R (probability of a response) to the same stimulus for both the monocular and binocular viewing conditions. There was some internal evidence in this study that the P_R to a given suprathreshold brightness was greater for a single brightness than for a method of limits determination while the P_R to a subthreshold brightness was lower for a single brightness than for a method of limits determination. If this were the case, since the binocular P_R will be higher than the monocular merely by the combination of independent probabilities, and since the brightness values were typically chosen so that the monocular and binocular P_R's straddle the threshold, the observed difference between monocular and binocular thresholds could have been inflated by this effect. A direct comparison of the single brightness with method of limits methods of threshold determination was not possible in this experiment.

The present experiment undertakes such a comparison. The amount of summation obtained by a single brightness procedure and the method of limits is compared.

In addition, two methods of pro-

[1] The experiment reported in this paper was performed at Duke University and was supported by a grant from the Duke University Council on Research.

ducing the viewing conditions, shutters and eyepatches, are compared to test the possibility that *S*'s knowledge of the viewing conditions may influence the results.

A third possibility considered is that the amount of summation observed in the previous experiment may occur only under conditions producing maximal sensitivity previously used. This is examined by repeating a portion of the previous experiment under more typical stimulating conditions.

Method

Apparatus.—The apparatus used in this experiment was a Model III Hecht-Shlaer Adaptometer. It presented every 4 sec. a 3° circular patch 7° below a red fixation point at a viewing distance of 25 cm. for a duration of .2 sec. The luminance of the patch was continuously variable from 1.5 log $\mu\mu$L to 5.4 log $\mu\mu$L. The color of the patch was determined by a filter having a maximum transmittance at 480 mμ. Calibration was maintained photometrically. Two shutters, separately operated by Synchron motors, were mounted between the headpiece and the stimulus patch. These shutters were located so that they could interrupt the visual patch to either eye separately. The fixation point was *always* viewed binocularly. The operation of the shutters was not detectable to *S*. The eyepatches used in a portion of the experiment were black plastic and were held in place by an elastic band. A head and chin rest were used. The *S*s responded by use of a key. The apparatus was housed in a sound-deadened, light-tight, ventilated chamber.

Subjects.—Six paid university students, three men and three women, served as *S*s. All met or exceeded naval visual standards as tested by the Orthorator (20/20 visual acuity, O.D., O.S.; no phorias; "color" normal). The *S*s had not had extensive previous experience in visual threshold determinations. They were instructed to maintain fixation and to push a lever to the right if they saw the flash, and to the left if they did not. At the beginning of each experimental period each *S* had spent at least 20 min. in dark adaptation glasses and an additional 20 min. in complete darkness. The *S*s participated in the experiment for 10 days with each session falling at the same time of day.

Training.—The six *S*s were divided, for purposes of training, into three groups of two each: R (monocular right), L (monocular left), and B (binocular). During training the stimulus patch was presented to the right eye only (R group), left eye only (L group), and both eyes (B group). The training sessions for all groups consisted of a standard ascending-descending ML (method of limits) procedure consisting of 30 alternately ascending and descending series. There were four training sessions, Days 1–4.

ML-shutters.—The first ML test series using the shutter consisted of two sessions, Days 5 and 6. Each session for each *S* consisted of six blocks of five ascending and descending series of stimulations. The B, R, L conditions were randomly assigned to these blocks, with the restrictions that no condition followed itself and that each condition occurred twice.

First SB.—The first single brightness test series consisted of two sessions, Days 7 and 8. Each session for each *S* consisted of six blocks of 50 stimulations at an invariant brightness. The brightness used was determined for each *S* from his preceding average R and L threshold to yield between 30% and 40% "yes" responses, under the monocular condition. The B, R, L conditions were randomly assigned as before to the blocks. The shutter method of occlusion was used. Up to this point only one *S* reported that he was aware of any changes in viewing conditions.

ML-eyepatch.—The second ML test series consisted of two sessions, Days 9 and 10, using the eyepatches. Six blocks of five ascending and descending series were used. The B, R, and L conditions were randomly assigned as before to these blocks. The eyepatch method of occlusion was used, with *S*'s being instructed to shift the eyepatches between the blocks of trials.

Second SB.—The second single brightness series, Days 11 and 12, using the shutter, consisted of two sessions of three blocks of 100 stimulations at an invariant threshold brightness. The B, R, and L conditions were assigned randomly. This series was run for only four of the six *S*s.

Results

For the ML data, measures of the threshold and dispersion of the "yeses" were obtained from a probability of response (P_R) plot. The two blocks of a given condition of a session were combined and percentage of "yes" was plotted against log brightness on normal probability paper. The 50% point (threshold)

and the 84%–50% (1/slope) difference were interpolated from the "visually" best fitting straight line.

For the SB data the percentage of "yes" responses for each block of trials was computed. To test for intertrial association a runs test (**3**) based on the expected number of runs at a given percentage of "yes" was employed.

Training sessions.—An analysis of variance on thresholds for training groups (B, R, and L) and sessions (Days 1–4) yielded a significant Session by Group interaction ($P < .001$; $F = 26.71$; $df = 6$ and 9). Both the B and L groups showed a significant decline in threshold, while Group R showed an increase (on Day 4). Though the mean threshold for Group B (Day 4 $Th_B = 2.51$ $\mu\mu$L) was less than that for Groups R and L (Day 4 $Th_L = 2.88$ $\mu\mu$L), the difference was not statistically significant.

An analysis of the 84%–50% differences for the same variables produced no significances.

Method of limits test sessions.—Table 1 presents the average threshold for training conditions, method of occlusion, and viewing conditions. An analysis of variance of these thresholds for Training conditions, R, L, or B; Viewing conditions, R, L, or B; Method of occlusion, shutter vs. eyepatch; and Sessions, first or second, is presented in Table 2. This analysis is an extension of Lindquist's type VI (**7**, p. 296). Both Viewing condition and Training condition proved to be significant. The average binocular threshold was 2.61 log $\mu\mu$L, while the average monocular threshold was 2.76 log $\mu\mu$L for both the right and left eyes, respectively. The average threshold, under all viewing conditions, for those *S*s trained binocularly was 2.53 $\mu\mu$L, while those *S*s trained on the right and left eyes were 2.91 and 2.69 log $\mu\mu$L, respectively. There was no interaction between Viewing condition and Training group. On the assumption that the two eyes operate independently, $P_B = P_R + P_L - P_R P_L$. Using the average threshold value for the monocular conditions, it would be expected that the P_B for these values would be 75%. The observed values were interpolated from the binocular P_R plots. For every one of the 24 possible comparisons, 6 *S*s by 4 days,

TABLE 1

METHOD OF LIMITS THRESHOLDS

Training	Occlusion	Viewing Conditions			Mean
		Binocular	Monocular Right	Monocular Left	
Binocular	Shutter	2.42	2.67	2.66	2.53
	Eyepatch	2.36	2.53	2.53	
Monocular right	Shutter	2.76	2.89	2.92	2.91
	Eyepatch	2.89	2.97	3.01	
Monocular left	Shutter	2.59	2.77	2.74	2.69
	Eyepatch	2.62	2.70	2.70	
Mean		2.61	2.76	2.76	

TABLE 2

ANALYSIS OF VARIANCE ON METHOD OF LIMITS THRESHOLDS

Source	*df*	*MS*	*F*
Between *Ss*	(5)	(.3874)	
Training group (G)	2	.8585	11.71*
Error (b)	3	.0733	
Within *Ss*	(66)	(.0140)	
Viewing condition (A)	2	.1892	32.62**
Mode of occlusion (B)	1	.0023	.24
Session (D)	1	.0415	6.59
A × B	2	.0098	1.29
A × G	4	.0075	1.29
A × D	2	.0027	.63
B × G	2	.0692	7.21
B × D	1	.0205	1.78
G × D	2	.0018	.29
A × B × G	4	.0004	.05
A × B × D	2	.0024	.53
A × G × D	4	.0054	1.26
B × G × D	2	.0119	1.03
A × B × G × D	4	.0037	.82
Error (w)	(33)	(.0065)	
$Error_1$ (w)	6	.0058	
$Error_2$ (w)	3	.0096	
$Error_3$ (w)	3	.0063	
$Error_4$ (w)	6	.0076	
$Error_5$ (w)	3	.0115	
$Error_6$ (w)	6	.0043	
$Error_7$ (w)	6	.0045	
Total	71		

Note.—The error terms for the within analysis follow the following pattern: A and AG against $Error_1$, B and BG against $Error_2$ . . ., AB and ABG against $Error_4$ . . ., and ABD and ABDG against $Error_7$.

* $P < .05$.

** $P < .01$.

the observed value exceeded the expected value. The mean excess was 19.0% for the shutter sessions and 15.2% for the eyepatch sessions. Thus, the binocular thresholds are significantly lower than the monocular and this difference is greater than that expected by the combination of independent probabilities.

When the same analysis, as presented in Table 2, is performed using an average of the monocular values, which results in a less variable estimate of the monocular values, the interaction between Viewing condition and Method of occlusion is significant. The binocular threshold is less for the shutters (2.59 log $\mu\mu$L) than the eyepatch (2.62 log $\mu\mu$L) while the average monocular threshold is greater for the shutter (2.77 log $\mu\mu$L) than the eyepatch (2.74 log $\mu\mu$L).

An analysis similar to that in Table 2 was performed on the 84%–50% differences. Only the interactions between Method of occlusion and Training group and between Method of occlusion, Training group, and Session reached significance: $P < .05$ ($F = 10.46$, $df = 2$ and 3) and $P < .01$ ($F = 24.75$, $df = 2$ and 3), respectively. This interaction appeared to result from increase in slope (smaller 85%–50% difference) of the P_R curve for the second session of each occlusion condition with the single exception of the L training group using the shutters. The slope differences were smaller for the eyepatch sessions, which had generally greater slopes.

Single brightness test sessions.—An analysis of variance, similar to that in Table 2, on Training group, Viewing condition, block of (50) Trials, and Sessions was performed on the percentage of "yes" responses of the first two SB sessions. Only Viewing condition reached significance: $P < .01$ ($F = 55.97$, $df = 2$ and 6). The average percentage of "yes" response for B was 74.7 while R and L were 30.5 and 28.5, respectively. Thus for the same brightness, the percentage of "yes" response was significantly greater on those occasions when the stimulus was viewed binocularly than when it was viewed monocularly. The expected value on the basis of the rule for combination of independent probabilities was calculated from the observed monocular probabilities. Again the observed value exceeded the expected for every one of the 24

possible comparisons. The mean excess was 26.0%.

In order to compare the data obtained by ML and by SB methods, the difference in percentages between the monocular and binocular SB data was converted to increments in brightness. Using the average SB monocular percentage of "yes," the increment in log brightness necessary to produce the observed increment in percentage of "yes" was computed from the binocular ML P_R curve. The average increment was .18 log $\mu\mu$L for ML and .23 for SB. These values were compared by analysis of variance. The difference between SB and ML was significant: $P < .05$ ($F = 18.00$, $df = 1$ and 3). Similar results were obtained for the second two sessions (blocks of 100 trials).

Squaring and cumulating the z scores obtained from the runs test (**3**) for each S for the first two SB sessions yielded a chi square with 12 df, which showed four of six Ss with a significant amount of intertrial association: $S_1 = 16.90$, $S_2 = 29.28$, $S_3 = 25.73$, $S_4 = 8.59$, $S_5 = 56.67$, and $S_6 = 23.15$; $P = .05$ ($\chi^2 = 21.03$, $df = 12$). Using the same procedure for the second two SB sessions, three of the four Ss showed a significant amount of intertrial association; $S_2 = 58.22$, $S_4 = 30.53$, $S_5 = 35.11$, and $S_6 = 10.15$ ($P = .05$, $\chi^2 = 12.59$, $df = 6$).

Analyses of variance over the variables of the experiment on the z scores computed from the runs test yielded no significant differences.

Discussion

In the first column of Table 3 are presented the monocular-binocular brightness from this experiment and from the preceding study. In the second column are the mean differences between the expected binocular percentage of "yes" and the obtained percentage. The results of the present study support those of the previous study on both the existence and order of magnitude of threshold binocular summation. Stimulation of two eyes seems functionally equivalent to doubling the area of stimulation in a single eye, as can be seen by comparison with the values expected from Piper's law or Crozier's areal function. This statement must be qualified in a number of ways.

TABLE 3

Comparisons of Monocular and Binocular Thresholds Differences

Condition	Mean Diff. Between Binocular and Av. Monocular Thresholds	Mean Diff. Between Expected and Obtained Binocular % "Yes"
ML shutter	.18 log $\mu\mu$L	19.0%
ML eyepatch	.12 log $\mu\mu$L	15.2%
SB shutter	.23 log $\mu\mu$L	26.0%
Previous study	.18 log $\mu\mu$L	14.9%
Piper's law	.15 log $\mu\mu$L	
Crozier's	.18 log $\mu\mu$L	

First, there is the problem of the appropriate relation between area of the stimulus and the threshold. For an exact test of binocular stimulation it is necessary to be able to predict the threshold decrease occasioned by doubling the area of stimulation for the stimulating conditions used. The relation has been determined under only a limited set of circumstances. It is subject to a variety of conditions such as shape, location, etc. Although the results of the present and previous experiments show that summation of the same order of magnitude occurs under a widely differing set of stimulating conditions, a direct comparison of the values from the point of view of the areal function is not possible without a direct measurement of this function under the two conditions. A recent study by Leibowitz and Walker (**6**), in which they found that field size was an important parameter in the amount of summation observed, further points out

the difficulty of making direct comparisons between different stimulating conditions.

Second, as demonstrated by the present experiment, thresholds, and therefore measures of summation, are subject to a number of procedural variables. The conjecture concerning the magnification of the summative effect by the single brightness procedure was confirmed by the present experiment. The SB procedure resulted in a significantly greater separation of the monocular and binocular thresholds. Further support for this interpretation of the differences between the two methods is given by some recent direct evidence (**2, 5**) that the P_R to a stimulus presented by the SB procedure is greater than that obtained by a procedure in which the successive stimuli vary in magnitude, if the P_R is greater than 50%; while the P_R to a stimulus presented by the SB procedure is less than that obtained by a varying stimulus procedure if the P_R is less than 50%. Bartlett and Gagne (**1**) have suggested that summation occurs as a result of the cumulative retinal changes produced by repetitive stimulation. The present results support the view that the monocular-binocular difference is enhanced by repetitive stimulation, i.e., the SB method, but it does not support repetitive stimulation as an explanation of summation. Although the difference between monocular and binocular thresholds is not an artifact of the method of threshold determination, its magnitude appears to be sensitive to it.

A second procedural variable which influenced the value of the thresholds obtained was training conditions. Initially, those *S*s who experienced four days of training in monocular viewing have higher thresholds under both monocular and binocular test conditions. The magnitude of the difference between the monocular and binocular thresholds does not seem to be influenced. The training differences disappear over days and fail to be significant in later sessions. Monocular and binocular viewing seem to be functionally different tasks; training on one does not necessarily facilitate performance on the other. The effect of training sessions on the slope, i.e., the initially steeper slope (and higher threshold) for the R group, suggests that eye dominance may play a part.

Any interpretation of the effect of the different modes of occlusion on the difference between monocular and binocular thresholds is made difficult by the impossibility, in the present experiment, of counterbalancing the order in which these conditions were given. Whether the outcome of the present experiment (*a*) is an artifact of the lack of counterbalancing, (*b*) is an artifact of the fixation conditions, or (*c*) represents some real differences between the conditions, cannot be evaluated. The first possibility could result from the improvement with practice of *S*s in the possibly more difficult monocular viewing conditions producing lower monocular thresholds in the later sessions and thus a smaller difference. The second possibility receives some support from the previous experiment in which there was a suggestion that the binocular (vs. monocular) *fixation* condition resulted in slightly higher monocular thresholds and, consequently, a larger difference between monocular and binocular thresholds. In this experiment the monocular fixation conditions (eyepatch procedure) again produced the lowest thresholds. The possibility that the difference is real remains.

It is clear from the present results that procedures of stimulation as well as the stimulating conditions have to be carefully considered when tests of hypotheses involve measurements of thresholds, as in the binocular summation problem. The fact that the same magnitude of summation was found under the somewhat less rigorous stimulating conditions in the present experiment as compared with the preceding study, suggests that the diversity of results found in experiments on summation may stem in large part from procedural differences rather than the differences in stimulating conditions.

The present findings with respect to

the magnitude of intertrial association, monocularly and binocularly, support the previous conclusion concerning the nonperipheral origin of the associative effect: fixation, pupillary changes, and eye movements do not influence it.

Summary

The method of limits and single brightness procedures of stimulus presentation were compared in their effect on binocular summation. Summation exceeding that expected by the combination of independent but nonmutually exclusive probabilities was found in both cases, but, as conjectured, the single brightness procedure produced the greatest difference between the monocular and binocular thresholds.

The magnitude of summation found was of the same order as that in a preceding experiment employing quite different stimulating conditions.

Both training conditions and mode of occlusion were found to affect the thresholds and the latter also affected the size of the difference between the monocular and binocular thresholds.

References

1. Bartlett, N. R., & Gagne, R. M. On binocular summation at threshold. *J. exp. Psychol.*, 1939, **25**, 91–99.
2. Brackman, J., & Collier, G. The dependence of probability of response on the size of the step interval in method of limits. *J. exp. Psychol.*, 1958, **55**, 423–428.
3. Collier, G. Intertrial association at the visual threshold as a function of intertrial interval. *J. exp. Psychol.*, 1954, **48**, 330–334.
4. Collier, G. Probability of response and intertrial association as functions of monocular and binocular stimulation. *J. exp. Psychol.*, 1954, **47**, 75–83.
5. Collier, G., & Verplanck, W. S. Nonindependence of successive responses at the visual threshold as a function of interpolated stimuli. *J. exp. Psychol.*, 1958, **55**, 429–437.
6. Liebowitz, H., & Walker L. Effect of field size and luminance on binocular summation of supra-threshold stimuli. *J. opt. Soc. Amer.*, 1956, **46**, 171–172.
7. Lindquist, E. F. *Design and analysis of experiments in psychology and education.* Boston: Houghton Mifflin, 1953.
8. Pirenne, M. H. Binocular and uniocular threshold of vision. *Nature*, 1943, **152**, 698–699.

(Received December 17, 1957)

Journal of Experimental Psychology
Vol. 56, No. 4, 1958

AMOUNT OF INFORMATION GAINED DURING BRIEF EXPOSURES OF NUMERALS AND COLORS[1]

NANCY S. ANDERSON[2] AND PAUL M. FITTS

Ohio State University

Questions relating to span of perception, or span of attention, have been asked since the early days of experimental psychology, and long before this similar problems were of interest to philosophers. The types of inquiries have ranged from those concerned with *S*'s ability to report how many stimulus objects were present during a brief exposure to those concerned with his ability to provide detailed descriptions of the objects. Still a different approach to this problem is taken in the present experiments, the specific question at issue in this instance being, "How much information can *S*s report after the tachistoscopic exposure of a group of symbols, as a function of the information content of the symbols and the method of information coding?" Such an analysis of perceptual capacity appears to be a logical extension of the trend, which began a number of years ago, to regard the present topic as one involving cognition rather than span of attention.

In 1929 Glanville and Dallenbach proposed that ". . . the constant of the traditional 'range of attention' experiment is dependent on the degree of cognition involved in the report" (**7**, p. 210), and demonstrated such effects experimentally. However, these investigators measured performance in terms of the conventional limen, defined as that value of stimulus which has a 50% probability of being correctly reported. This criterion measure, which has been employed by most previous workers, now appears to be somewhat inappropriate. In the first place, the traditional computation of a limen treats all errors alike, whereas numerous recent experiments, such as that of Bricker and Chapanis (**3**), have shown that information can be recovered by examining the frequencies of different types of errors. In the second place, the types of functions obtained for limens and for information-transmitted measures lead to quite different interpretations. Whereas use of the conventional limen always indicates a reduction in perceptual performance with any increase in the level of cognition required in a task, information analysis often shows first an increase, then a leveling off, and finally a decline in performance (information transmitted) as stimulus complexity or stimulus rate is increased (**1, 6, 8, 10**). Thus, in tasks involving different sizes of alphabets and different types of coding, the use of informational analysis enables *E* to determine optimal performance conditions. It appears, therefore, that the problem of "level of cognition" can be restated as the problem of "amount of information," and a more sensitive measure of performance substituted for the conventional limen.

[1] This research was conducted in the Ohio State University Laboratory of Aviation Psychology, and was supported in part by Contract No. Nonr-495(02), between the Office of Naval Research and the Ohio State University Research Foundation.

[2] Now at the International Business Machines Research Laboratory, Yorktown Heights, New York.

Recent evidence indicates that performance in processing information is a function of the number of alternative symbols in the coding alphabet. Alluisi, Muller, and Fitts (**2**) reported that the rate of information transmission in speeded perceptual-motor tasks increases as alphabet size is increased. Miller (**9, 10**), in summarizing recent studies of absolute discrimination, memory span, and span of perception, concluded that the amount of information transmitted often is improved by increasing the uncertainty per symbol, i.e., the size of the coding alphabet.

Since *S*'s ability to identify stimuli on an absolute basis is limited, the maximum number of stimulus categories that can be used when the stimulus is varied along only a single dimension is small. Therefore, the most feasible means for increasing the size of a coding alphabet is to increase the dimensionality of the stimulus. In the present experiment, three levels of stimulus dimensionality were studied. Alphabets of symbols consisted of (*a*) color patches, (*b*) black two-dimensional forms (numerals), and (*c*) black forms on colored backgrounds (color-numeric symbols).

Experiment I

Method

Symbols.—Nine homogeneous color patches formed one coding alphabet. Nine black numerals, "0" through "9" exclusive of "1", formed another alphabet. In neither of these alphabets was a strictly one-dimensional coding scheme employed, since the color matches varied somewhat in brightness and saturation, as well as in hue, and since form is certainly multidimensional. A third and still more complex type of alphabet was formed by superimposing black numerals on color patches. This particular color-numeric alphabet resembled somewhat the one studied earlier by Glanville and Dallenbach (**7**). It was selected after the tryout of numerous other schemes, such as colored numerals.[3]

Numerals were 1.2 cm. wide by 2 cm. high, with 3-mm. stroke width, and were of a style providing good legibility. Color patches were 2.5 × 3.2 cm. in size and were made of Color Aid Company papers. The nine colors were selected to be equally discriminable by extrapolating data published by Chapanis and Halsey (**4**) and were checked in a pilot study. The decision in regard to the maximum number of color categories to be used (nine) was based on the original data of Chapanis and Halsey (**4**) for spectral colors, and Conover's (**5**) data for colored papers.

Messages.—Messages were formed by selecting a group of symbols at random and with replacement (by means of random number tables) from one of the alphabets. In the case of the first two alphabets, color patches or numerals, four message lengths were used: 3, 4, 5, and 6 symbols. Since the choice of each symbol (from among nine equally likely alternative symbols) provides approximately 3.17 bits, the information contained in these messages varied from 9.51 to 19.02 bits per message.

The length of messages was held constant at three symbols in the case of the color-numeric alphabet, but information per symbol was varied by successive restrictions on the size of the set of alternative color symbols used. In one case, all nine possible colors were used in all combinations with the nine numerals, so that the choice of each complex (color-numeric) symbol provided $\log_2 81$, or 6.34 bits, and a message three symbols in length contained 19.02 bits. In other cases, however, only six and three possible colors were used in combination with the nine numerals, reducing the information, respectively, to $\log_2 64$ and $\log_2 27$ bits per complex symbol, and the total information per three-symbol message to 17.25 and 14.25 bits, respectively.

Exposure apparatus.—Symbols were exposed for .1 sec. behind a double-bladed shutter which was operated by a rotary solenoid. This exposure apparatus was mounted behind a 15 × 4-cm. opening in a large gray screen. The level of direct illumination on the symbols was 38 ft-c.

Subjects.—Nine men and three women, all volunteers, served as *S*s.

Procedure.—The *S*s were tested in groups of four. Two *S*s sat at floor level 10 ft. from the screen; the other two sat on a 4-in.-high platform 12 ft. from the screen. Each *S* was given one practice session and 10 test sessions. During the training session each *S* received 18 trials with each

[3] This preliminary work was carried out by Raymond C. Sidorsky in collaboration with the second author of the present report.

of the three types of symbols, and with messages containing each of the different levels of information.

The names assigned to the colors were Red, Yellow, Blue, Green, Violet, Flesh, Orange, Pink, and Indigo, in order that each name would have a unique first letter. Initial paired-associate training was given to ensure overlearning of these color names, and refresher training was given before each test session.

During the first five test sessions data were collected only for colors and numerals. Each session consisted of 36 test messages for each type of symbol and for each of the four message lengths. Each group of four *Ss* worked under these conditions in a different random order. The complex symbols were used on the last five test sessions.

Information was given as to the type of messages that would be exposed and the size of the set of alternative symbols that would be used in each series of trials. An alerting signal sounded 3 sec. before each message exposure, and *Ss* were instructed to fixate a dot on the center of the shutter as soon as they heard this warning. The *Ss* recorded first the color symbol and then the numeral for each complex color-numeric symbol (this sequence, of course, was arbitrary). They were told to guess, even if they had no idea what they had seen, i.e., to fill in all response blanks.

TABLE 1

AVERAGE INFORMATION GAINED IN BITS FOR THE TWO SIMPLEST ALPHABETS IN EXP. I
(Averages based on data computed separately for each of 12 *Ss*, 180 responses per *S* per condition)

Number of Symbols per Message*	Symbol Position (Left to Right)						Mean per Symbol
	1	2	3	4	5	6	
Numerals							
3	3.15	3.14	3.14				3.14
4	3.15	3.15	3.14	3.15			3.15
5	3.12	3.08	3.05	2.71	2.81		2.95
6	2.83	2.92	2.84	2.16	1.65	1.92	2.39
Colors							
3	3.02	2.91	2.82				2.91
4	2.84	2.61	2.30	2.27			2.50
5	2.29	2.04	1.68	1.10	0.90		1.60
6	2.04	1.99	1.44	1.11	0.59	0.52	1.28

* H_o per stimulus symbol = 3.17 bits.

Results

The responses made by each of the 12 *Ss* were analyzed separately for each of the symbol positions in each type of message and for each type of coding. Learning effects, after the initial training session, were very small and therefore the data for the same *Ss* under identical conditions were pooled across sessions.

The results are summarized in Tables 1 and 2. Data are summarized separately for each symbol position, for each message length, and for colors and numerals separately in the case of the color-numeric symbols. The total amount of information transmitted in reporting each type of message is shown in Fig. 1 (which also shows some results from Exp. II).

Performance with the complex color-numeric alphabet was greatly superior to performance with either colors or shapes alone. The average amount of information transmitted following the exposure of three complex symbols was 16.97 bits. This was significantly greater ($P < .01$ for $N = 12$ *Ss*) than the amount transmitted (14.30 bits) for six numerals, and this in turn was significantly better than performance with six color patches (7.69 bits). Differences among the three types of messages increased as the total information in the message increased.

There is an indication in the data that performance reaches a peak and then declines as message length is increased. Performance with colors was significantly better for messages containing only four symbols than for longer messages. Performance with numerals was slightly better when *Ss* saw five rather than six symbols, but the drop in performance was not statistically significant. Mean performance with the complex symbols had not reached a peak for the highest information messages used.

TABLE 2

AVERAGE INFORMATION GAINED IN BITS FOR THE COMPLEX ALPHABET IN EXP. I
(Averages based on data computed separately for each of 12 *Ss*,
180 responses per *S* per position)

Stimulus Information per Symbol Position			Symbol Position (Left to Right)						Mean per Symbol	
			1		2		3			
Color	Shape	Sum	Color	Shape	Color	Shape	Color	Shape	Color	Shape
1.58	3.17	4.75	1.54	3.10	1.50	3.09	1.53	3.07	1.52	3.08
2.38	3.17	5.75	2.45	3.04	2.36	2.99	2.23	2.99	2.34	3.01
3.17	3.17	6.34	2.88	2.91	2.84	2.92	2.64	2.82	2.77	2.88

For the first two alphabets the most information, on the average, was transmitted regarding the symbol in the left-hand position (which was recorded first) and the least information was transmitted regarding the last two symbols on the right. These effects were largest for the longest messages. Individual *Ss* varied considerably in regard to position effects; a few did best on the center positions.

When three color-numeric symbols were used, each of the symbol positions conveyed an approximately equal amount of information, and *Ss* reported almost as much information about color as they did about shape.

EXPERIMENT II

Because performance with the color-numeric symbols was found to be highest for the maximum information messages employed in Exp. I, whereas performance in reporting messages of similar information content but comprised of less complex symbols had apparently passed a peak and begun to decrease, it was decided to conduct a further experiment including messages of higher information content. Some of the lower information conditions of Exp. I were repeated for comparison purposes.

Method

Symbols.—The first two alphabets were the same as those employed in Exp. I; the third alphabet was identical to the most complex color-numeric alphabet used in Exp. I, i.e., nine numerals superimposed on nine color patches.

Messages.—Messages of three levels of information content, 12.68, 19.02, and 25.36 bits per message, were used. These were formed by using 4, 6, or 8 simple symbols or by using 2, 3, or 4 combined symbols per message.

Subjects.—Twelve volunteer students served as *Ss*. None had served in the earlier experiment.

Procedure.—All procedures were identical to those followed in Exp. I.

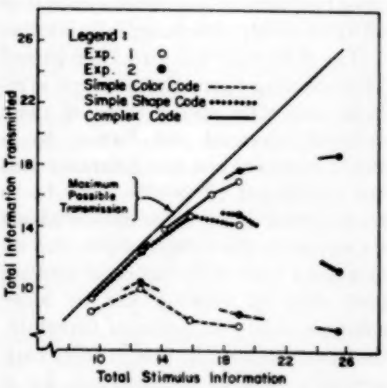

FIG. 1. Mean total information gained (transmitted) per message by *Ss* in Exp. I and II (N = 12 different *Ss* in each experiment).

Results

Performance as a function of symbol position within a message followed the same general pattern as in Exp. I, and only mean performance per message need be considered here. These data are shown in Fig. 1.

Six conditions were identical in Exp. I and II, except that new *Ss* were used. The relative magnitudes of these six points are identical in the two studies, but all points in Exp. II are slightly higher (by about .1 bit per simple symbol and .2 bit per complex symbol) than corresponding points in Exp. I. These results provide strong confirmation of the findings for Exp. I. The slight but consistent superiority of *Ss* in Exp. II could easily be due to sampling of individuals, since the range of individual differences among *Ss* within groups was of the order of 1 bit per symbol position in a message.

The use of color-numeric symbols again led to significantly better performance than did the use of colors or numerals alone ($P < .01$). All 12 *Ss* did better in reporting about the more complex symbols than in reporting about either of the simpler symbols at all three levels of message information.

The difference in information gained after viewing four color-numeric symbols, and after seeing three of these symbols, favored the former by a slight amount, but the difference was not significant ($P > .05$). All 12 *Ss* transmitted more information about six numerals than about eight, and all *Ss* gained more information in viewing four than in viewing five or more color patches. It appears, therefore, that performance in this type of task generally reaches an optimum for a given type of coding, as stimulus information is increased, and then declines.

Critical Evaluation of the Present Use of Information Analysis

The discussion of the present results is in two parts. The first deals with the limitations of the information measure when employed in this type of study, and considers some of the factors affecting the amount of information conveyed by absolute judgments (see also the recent discussion by Alluisi [1]). Such a critical evaluation is indicated in view of our previously stated criticism of the traditional methodology involving the computation of limens.

Sequence followed in recording responses.—In order to check the supposition that the symbol dimension which was written down first was favored, the two possible orders of reporting color and shape were investigated in a supplemental study. The dimension recorded first was found to be reported significantly more accurately. Thus, in Exp. I and II the superiority of shape over color would probably have been even greater had the arbitrary order of reporting been reversed.

Central vs. peripheral vision.—Part of the relative superiority of complex over simple symbols may be due to the fact that a greater proportion of the complex symbols were viewed in central vision since they can be arranged more compactly for a given amount of information per message. In order to check the supposition of a central vs. peripheral vision effect, performance with the complex symbols was also determined for a condition in which they were grouped in a square pattern, resembling the "square span" typographic arrangement studied by Nahinsky (11), and it was found that significantly fewer errors were made in reporting about six symbols arranged in a square as compared with a line pattern. It is concluded that their more compact visual pattern accounted for part of the superiority of messages comprised of the complex symbols, and that information transmission per tachistoscopically exposed message can be increased by compact grouping of symbols. It should be noted that this is a perceptual effect, since the amount of information to be remembered and the mode of subsequent responding was the same regardless of the grouping of the stimulus symbols.

Nonindependence of errors.—In the present analysis each symbol position (and dimension) in a message was treated separately. This was a necessary expediency, since by this procedure each set of pooled data could be summarized in a 9×9 contingency table with cell frequencies that were sufficiently large for the $p \log p$ operator employed in information analysis. Such an independent analysis of each symbol position,

however, does not correct for any dependencies among the responses to successive symbol positions. The computations therefore provide an underestimate of transmitted information whenever additional information could be recovered by analyzing all possible relations among errors. The alternative is to analyze as a unit the entire series of *S*s' responses; however, this would have required, in the case of the maximum message in Exp. II, the compilation of frequency data for a contingency table having almost a million entries for each axis, clearly an impractical undertaking. One check was made on the degree of underestimation introduced by the present computational simplification. One of the most commonly expected effects, if errors are correlated, would be a response in which adjacent pairs of symbols are reversed (a transposition error). The frequency of this particular type of error was checked by hand in a pilot study and found to occur with only a very slightly greater than chance frequency, a frequency which would have a negligible effect on the estimate of information transmitted. The extent of the underestimation of information transmitted because of the neglect of transposition errors and similar effects would therefore appear to be slight, and in any event should be relatively uniform for all experimental conditions.

Individual vs. group data.—In the present study a separate information analysis was made on the data for each of the 12 *S*s. Some *E*s have preferred to pool their data before applying the information analysis, since this also simplifies the computations. To the extent that *S*s make consistent and peculiar types of errors not ordinarily made by other *S*s, however, the present procedure provides the correct estimate of information transmission capacities of individuals. It also permits statistical tests of the significance of mean differences, based on variance among *S*s. Estimates of average information transmission computed in this manner are slightly higher than estimates computed from pooled data.

*Level of training of S*s.—It has been shown (**12, 13**) that slow but significant improvement in performance results from continued tachistoscopic training. The present data, therefore, should not be interpreted as indicating the maximum abilities of *S*s to learn to extract information from symbols during a brief exposure period. Instead, the results should be generalized to conditions where only moderate amounts of training are given.

Discussion

The present experimental findings indicate that the amount of information that can be reported about stimuli exposed for .1 sec. reaches a maximum for some particular message length and complexity, and becomes less for longer or more complex messages. Maxima occurred at different levels of stimulus information for the three alphabets examined in the present experiments. Since similar types of functions have been found in studies of information handling rate in serial tasks, it appears that performance in both discrete and in serial tasks is optimum for some specified range of stimulus information. Whereas the traditional question regarding perceptual ability has been "How many stimulus objects can *S* count, identify, or describe correctly?" it appears that a more useful form of the question is "Under what conditions is the amount of information that *S* can report about a group of stimulus objects at a maximum?"

It is obvious from the present results that optimum performance, as measured by the amount of information transmitted per message, is a function of the stimulus code employed. One factor which limits performance in the present task is the ability of *S*s to identify each symbol on an absolute basis. Although most *S*s can identify the nine symbols employed with the color alphabet under normal viewing conditions, it is possible that they experience difficulty in discriminating each of the colors, and perhaps even in discriminating some of the shapes, under tachistoscopic-exposure conditions. Another limiting factor is memory ability, including such recoding (naming) of symbols as is required before they can be reported. On the basis of these two considerations it would appear that the information gained during a brief exposure of a coded message should be maximum when the number of symbols is within the memory span of *S*s, and when some form of multidimensional coding is employed which does not exceed *S*s' ability for absolute discrimination. In theory, under these conditions, the limit on amount of information gained should be set by *S*s' facility in recoding the information conveyed by the com-

plex symbols into a form suitable for short-term memory.

The use of shape-coded symbols, which in some respects are more complex than color symbols, permitted higher maximum information transmission than did color symbols, and the use of combined color-numeric symbols enabled *S*s to report an even greater maximum amount of information. Thus, in these three instances the greater the complexity or "dimensionality" of the symbols comprising an alphabet, the greater was the peak amount of information that was reported about the displayed message. This finding agrees with the preceding argument and also is in agreement with results obtained in other types of tasks, such as sorting objects (**6**), extracting information from dot and matrix patterns (**8**), and responding quickly to numeric symbols drawn from alphabets of differing size (**2**). However, the total information gained even with the most favorable of the present codes is not particularly impressive. Independent studies indicate that approximately seven categories can be discriminated along most simple stimulus coding dimensions, and that about seven "chunks" of information can be remembered momentarily (**10**). If *S*s could discriminate as well when viewing seven stimuli simultaneously, and remember accurately about all seven objects, the net amount of information gained would be approximately 20 bits ($7 \log_2 7$). It is perhaps only fortuitous that maximum performance in the present task was at approximately this level.

Use of the background space surrounding a form to convey information, along with the information conveyed by the form itself, may have numerous practical applications, such as in situations where the total area available for a display is limited. Also, as pointed out earlier, multidimensional coding permits more efficient use of central vision. Theoretically, the ability of *S*s to observe and report approximately equal amounts of information regarding both figure and background characteristics, and to report more total information in such cases than when they report only about one or the other aspects of the stimulus field, supports the generally held view that perception of figure and ground is a unitary process.

Summary

In the present experiments the problem of how much *S* can perceive during the brief exposure of a group of stimulus objects is re-examined from the viewpoint of the relation of amount of information transmitted to information coding and amount of information displayed.

Groups of homogeneous color or shape symbols, or combined color-numeric symbols, were exposed for .1 sec. The information content of the groups of symbols (messages) varied from 9.51 to 25.36 bits per message. The *S*s wrote down estimates of all the symbols in each message. Results were analyzed separately for each *S* in terms of information transmitted.

Performance at first increased and then decreased as information content per message was systematically increased, reaching different maxima for each of the three coding schemes studied. From the present viewpoint, therefore, it is not meaningful to summarize perceptual ability in terms of the number of objects that can be reported following a brief exposure; it can only be said that reporting ability is maximum for messages of some specific level of complexity (information content).

Maximum average information transmitted about color (Exp. II) was 10.44 bits per exposure, with messages that contained four color symbols. Maximum average information transmitted about numerals (Exp. II) was 14.94 bits, with messages that contained six symbols. Maximum information transmitted about the color-numeric symbols was 18.64 bits, with four symbols; this was significantly greater than the information transmitted after exposure of any of the messages comprised of the two simpler alphabets.

In spite of the relative superiority of the color-numeric alphabet, the information gained from viewing groups of these complex symbols for a .1-sec. interval is somewhat less than would be predicted from independent data regarding (*a*) ability to identify stimulus levels along single coding dimensions on an absolute basis under favorable viewing conditions, and (*b*) immediate memory ability.

The superiority of the color-numeric symbols, whose use required *S*s to report data regarding both figure and background, is consistent with the notion that figure-ground perception is a unitary process; however, this superiority may

be due at least in part to more effective use of central vision and/or to more effective recoding of information for immediate memory.

References

1. Alluisi, E. A. Conditions affecting the amount of information in absolute judgments. *Psychol. Rev.*, 1957, **64**, 97–103.
2. Alluisi, E. A., Muller, P. F., Jr., & Fitts, P. M. An information analysis of verbal and motor responses in a forced-paced serial task. *J. exp. Psychol.*, 1957, **53**, 153–158.
3. Bricker, P. D., & Chapanis, A. Do incorrectly perceived stimuli convey some information? *Psychol. Rev.*, 1953, **60**, 181–188.
4. Chapanis, A., & Halsey, R. M. Absolute judgments of spectrum colors. *J. Psychol.*, 1956, **42**, 99–103.
5. Conover, D. W. The amount of information in the absolute judgment of Munsell hues. Unpublished doctoral dissertation, Ohio State Univer., 1955.
6. Eriksen, C. W. Object location in a complex perceptual field. *J. exp. Psychol.*, 1953, **45**, 126–132.
7. Glanville, A. D., & Dallenbach, K. M. The range of attention. *Amer. J. Psychol.*, 1929, **41**, 207–236.
8. Klemmer, E. T., & Frick, F. C. Assimilation of information from dot and matrix patterns. *J. exp. Psychol.*, 1953, **45**, 15–19.
9. Miller, G. A. Information and memory. *Sci. Amer.*, 1956, **195** (2), 42–57.
10. Miller, G. A. The magical number seven, plus or minus two: Some limitations on our capacity for processing information. *Psychol. Rev.*, 1956, **63**, 81–97.
11. Nahinsky, I. D. The influence of certain typographical arrangements upon span of visual comprehension. *J. appl. Psychol.*, 1956, **40**, 37–39.
12. Paterson, D. G., & Tinker, M. A. Effect of training on perceptual span. *Amer. J. Psychol.*, 1947, **60**, 388–396.
13. Renshaw, S. The visual perception and reproduction of forms by tachistoscopic methods. *J. Psychol.*, 1945, **20**, 217–232.

(Received December 13, 1957)

Journal of Experimental Psychology
Vol. 56, No. 4, 1958

DISCREPANCY FROM ADAPTATION LEVEL AS A SOURCE OF AFFECT[1]

RALPH NORMAN HABER

Yale University

This study is an attempt to specify the antecedent conditions which produce affective arousal. Little theorizing or research has been done in this area even though the concepts of "affect" and "affective arousal" are being included in contemporary theories of motivation (**2, 7, 10, 11**). Whenever such a concept is employed, the theorist is pressed to find an antecedent definition of it, since in most cases a consequent defined concept of affect leads to circularity. McClelland, Atkinson, Clark, and Lowell (**7**) have developed a theory based, in part, on Hebb's (**5**) neurological model of the origin of affect and Helson's (**6**) notion of adaptation level. According to McClelland's discrepancy hypothesis as to the origins of affect, "Positive affect is the result of smaller discrepancies of a sensory or perceptual event from the adaptation level of the organism; negative affect is the result of larger discrepancies" (**7**, p. 43). The discrepancy hypothesis predicts that affects as a function of discrepancy from the adaptation level (AL) should be a symmetrical "butterfly" curve, as shown in Fig. 1. Affect is now stated to be a function, not of increasing physical intensity per se, but of the size of the discrepancy between the adaptation level (expectation or adaptation) of the organism, and the stimulus (perception).

The concept of adaptation level was developed by Helson. He regarded the AL as being experientially determined: "The effects of stimulation form a (central) spatio-temporal configuration in which order prevails." Within this pattern, "there is assumed a stimulus that represents the pooled effects of all the stimuli, and to which *S* is attuned or adapted" (**6**, p. 2).

Hebb's (**5**) neurological model of the origins of affect holds, briefly, that when nerve impulses reach the cortex, they set up cell assemblies; patterns of excitation in specific nerve nets that follow designated paths and orders of firing. After the assemblies have been established, according to Hebb, further excitation may have several effects. If the incoming impulses are in the same pattern as the established assembly, the assembly short-circuits, i.e., the excitation of any one element in the assembly will fire the whole pattern. This corresponds to boredom—to complete and exact fulfillment of all "expectations." If, however, these impulses are sufficiently different from the assembly to alter it slightly or to add new patterns to it, the conscious experience is pleasure or positive affect. The final possibility is excitation so different as to completely disrupt the assembly, resulting in displeasure or pain.

The present experiment is designed to test the discrepancy hypothesis as an explanation of variation in affect experienced in the "temperature sense." The adaptation level is that temperature of water to which *S*'s hands are adapted, and discrepancies are different temperatures of water in which *S*'s hands are put. The *S* is asked to judge his affective reaction to the temperature changes by comparing simultaneously two discrepancies from the same AL, one for each hand.

Adaptation occurs immediately to

[1] This research was performed at Wesleyan University while the author was on a Behavioral Science Fellowship from the Ford Foundation, and also with the assistance of a Grant-in-Aid from the Ford Foundation to David C. McClelland. The author gratefully acknowledges this double debt to the Ford Foundation and especially to Dr. McClelland. Portions of this paper were read at the 1958 APA convention.

temperatures identical with the normal skin temperatures of *S*, and in a matter of seconds to temperatures slightly above or below normal skin temperature, where a slight time lag is required for the skin temperature to shift to this new value. When the adapting temperature is increased sufficiently above or below skin temperature so that adaptation is impossible, then no AL can be established because the adapting temperature will feel warm or cool itself. Discrepancies beyond this point should no longer produce the relationship postulated in Fig. 1. The exact nature of the reaction to them cannot be predicted in advance, except that for very high temperatures it should be unpleasant to practically all further increases in temperature and for very low temperatures it should be unpleasant for all further decreases in temperature.

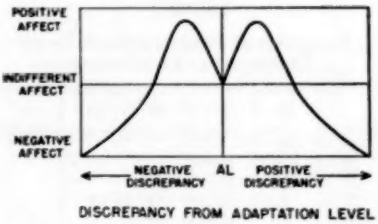

FIG. 1. Affect as a function of discrepancies above and below the AL (hypothetical).

METHOD

Apparatus.—The comparison stimuli were ten 10-qt. buckets of water arranged in a horseshoe, with an eleventh bucket as the standard stimulus, designated the AL bucket. The *S* sat in the center of the horseshoe, which was small enough to allow him to put both hands in all possible combinations of buckets. Heat was provided by 11 600-w. heating units, each with a variable resistor, and each connected to a variable transformer, having a maximum output of 7.5 amp. and 135 v. Water temperatures could be held constant over time by making heat input equal heat loss from the body of water. Adjustment and control of temperatures was possible to a degree finer than necessary for the discriminations made in this experiment. The size of the buckets was sufficient to prevent any shifts in temperatures resulting from immersion of the hand of *S* for short periods of time. All temperatures are stated in degrees Centigrade.

Procedure.—The *S*s were eight paid male Wesleyan University undergraduates. They were tested in individual sessions, each varying from 1 to 2 hr. In each session there were 80 to 135 judgments. As *S* entered the horseshoe for each session, the skin temperature of his hand was recorded by a chemical thermometer on the palm. Then he was told which bucket was the standard or AL bucket and was told to place both hands in it to the wrists. A light mark on the wrist was used to assure uniformity of immersion. The *S* was instructed to report when the water in the AL bucket felt neutral (i.e., no temperature sensation) to both hands. This was to obtain initial adaptation. Then *S* was instructed to place each hand simultaneously in a different comparison bucket, the correct ones being designated each time by *E*. The *S* then had to remove one of his two hands, the one that felt less pleasant. His report was required within 3 sec. after immersion, and no indifferent report was allowed. Then both hands were returned to the AL bucket and the process was repeated. It was emphasized by *E* that there was no "gimmick" or game to outguess, but that each choice was to be made on the basis of the pleasantness or unpleasantness of the individual comparison.

The temperature of the AL bucket was designated as 0° discrepant from the AL. Then all the comparison buckets were also designated as a specific discrepancy from the AL. Hence, if the temperature of the AL bucket was 33° and the two comparison buckets were 34° and 36°, then *S* had to make a hedonic choice between a discrepancy of +1° and +3°. If the AL was 34° and the two comparison buckets were the same as above, then the choice involved a 0° versus a +2° discrepancy. Table 1 shows the frequency of presentation of each comparison.

The Weber fraction for temperature sensation yields a j.n.d. of less than .1° in either direction (4) when the water is at skin temperature. This j.n.d. increases rapidly so that it is over 2.9° when the temperature is 48°. In view of this, the discrepancies chosen for the study are far enough apart to be easily discriminated.

In the climate of central Connecticut in spring, the average skin temperature is about 33°. However, extra buckets were maintained in the horseshoe to insure proper discrepancies

TABLE 1

FREQUENCY OF PRESENTATION OF EACH COMPARISON IN EACH SESSION

	0°	1°	2°	3°	7°
1°	15				
2°	15*	15*			
3°	15	15	15*		
7°	5	5	5*	5	
15°	5	5	5*	5	5

* In some sessions, a 2° discrepancy was not used. Its exclusion leaves a logarithmic scale of discrepancies.

in terms of the skin temperature of each *S* as he began the session. This was necessary when *S*'s skin temperature differed by more than a degree from 33°.

Three separate functions of the relationship between affect and discrepancy from AL were studied. (*a*) Five *S*s were used to establish the function when the AL was set at or near skin temperature and the comparison buckets *increased* by discrepancies of 0°, 1°, 3°, 7°, and 15°. Each *S* served in two sessions with these settings. (*b*) The same five *S*s were used to establish the function when the AL was set at or near skin temperature and the comparison buckets *decreased* by discrepancies of 0°, 1°, 3°, 7°, and 15°. Each *S* served in two sessions with these settings. (*c*) Three *S*s were used to study the effects on the butterfly function when the AL was *increased* above skin temperature. Each of these *S*s completed two sessions for each setting of the AL: 33°, 34°, 35°, 36°, 38°, and 40°, or 12 sessions for each *S*. In every one, the comparison buckets increased by discrepancies of 0°, 1°, 2°, 3°, 7°, and 15° from the AL. (There were no equivalent sessions for systematic decreases in the AL.) Note that for these three *S*s, when the AL was set near their skin temperature (one of the 6 AL's used), the same function was being tested as the five *S*s used in Part *a*. Hence, data relevant to the AL set near skin temperature will be presented for eight *S*s.

The data were paired-comparison judgments, which were transformed into a continuous scale of preferences on the basis of the number of times a discrepancy was preferred in relation to the number of times it was offered. Hence, if a particular discrepancy was offered 55 times, and in these it was preferred 40 times, it had a value of 73%. If some other discrepancy was offered 55 times and preferred 31 times, it had a value of 56%. The significance of the difference between any two values was tested by a χ^2 technique. For example, to test the difference between the 0 and the 1 discrepancies for a given *S*, a 2×2 χ^2 table was used, with marginals labeled: discrepancy, 1 and 2; and preference; preferred to all others; not preferred to all others. The judgments of preference between the 0° and 1° were split, half being included for the 0° and half for the 1°. This avoided double counting and inflated *N*'s.

All tests of significance were applied to one *S* at a time, and used data of only one session. For each *S* in each session, two χ^2 tests were computed, testing the highest preferred discrepancy against the two adjacent discrepancies. Hence, if the 1° discrepancy was preferred the most, it was compared to discrepancies of 0° and 3°.

The presentation order of the stimuli was randomly varied from session to session and from *S* to *S*, as was the position of the stimuli in the horseshoe. The buckets were so arranged that there were at least two buckets for each discrepancy stimulus, one on either side of the standard bucket. This prevented *S*s from learning place cues as quickly, and allowed the same comparisons to be tested using different

TABLE 2

AFFECT AS A FUNCTION OF THE DISCREPANCY OF TEMPERATURES ABOVE AND BELOW THE ADAPTATION LEVEL

(Cell entries are the percentages of times each discrepancy was chosen in relation to the number of times it was offered as a choice with all other discrepancies.)

S	Discrepancy					
	0°	1°	2°	3°	7°	15°
Above the Adaptation Level						
S_1	64	96		46	6	0
S_2	49	100		57	6	0
S_3	68	98		40	6	0
S_4	38	69		94	6	0
S_5	34	70		94	12	0
S_6	60	94	74	13	4	0
S_7	40	85	72	30	4	0
S_8	14	44	74	98	4	0
Below the Adaptation Level						
S_1	62	94		48	8	0
S_2	62	100		44	6	0
S_3	68	100		38	6	0
S_4	66	98		42	6	0
S_5	72	94		40	6	0

pairs of buckets. Consistent hand preferences were looked for by interspersing choices between equal stimuli. No *S* had to be eliminated on this account. The *S* never had to make judgments between stimuli where one was above and the other below the AL. Stimuli above the AL were always in a different session from stimuli below the AL. This was done to avoid any difficulties arising out of the possibilities of two distinct receptor systems being in operation, one for cold and another for hot.

Results

When the AL was set at or near the skin temperature for each *S* and the comparison buckets increased by a logarithmic scale (discrepancies of 0°, 1°, 3°, 7°, and 15° with the inclusion of a discrepancy of 2° for three *S*s), affect first increased and then decreased. This was found with all eight *S*s. For five *S*s, the maximum pleasure point (MPP) was a +1° discrepancy. For each of these *S*s, +1° was significantly ($P < .01$) preferred to either 0° or the next highest discrepancy. Three *S*s preferred a +3° discrepancy as the MPP. Analogous χ^2's reached at least the same significance levels. For each *S* in every session, there was only one maximum for the function, with a decrease on both sides of the maximum point.

When discrepancies less than the AL were tested, all five *S*s preferred an MPP of −1° discrepancy below the AL. All 20 χ^2 tests were significant ($P < .01$). Hence, for the five *S*s tested for both wings of the "butterfly," affect was symmetrically related to the discrepancies in each direction from the AL (see Table 2 and Fig. 2).

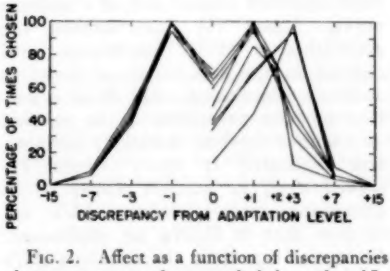

FIG. 2. Affect as a function of discrepancies of temperatures above and below the AL. Data taken from Table 2.

TABLE 3

AFFECT AS A FUNCTION OF DISCREPANCY OF TEMPERATURES ABOVE THE ADAPTATION LEVEL, WHEN ADAPTATION LEVEL WAS INCREASED FROM 33° TO 40°

(Cell entries are the percentages of times each discrepancy was chosen in relation to the number of times it was offered as a choice with all other discrepancies.)

S	AL	Discrepancy					
		0°	1°	2°	3°	7°	15°
S_6	33°	60	94*	74	13	4	0
	34°	74	91*	47	13	4	0
	35°	73	88*	48	13	4	0
	36°	97*	73	43	13	4	0
	38°	97*	72	43	13	4	0
	40°	100*	71	42	13	4	0
S_7	33°	40	85*	72	30	4	0
	34°	63	92*	61	21	4	0
	35°	59	99*	56	13	4	0
	36°	84	86*	43	13	4	0
	38°	98*	72	44	13	4	0
	40°	100*	71	42	13	4	0
S_8	33°	14	44	74	98*	4	0
	34°	15	55	70	86*	4	0
	35°	21	46	75	85*	4	0
	36°	56	67	75*	27	4	0
	38°	60	71*	61	35	4	0
	40°	100*	71	42	13	4	0

* Indicates the most preferred discrepancy.

In studying the effect of increasing the value of the AL with all three *S*s, the AL could be increased from 33° to 35° without changing *S*s' preferred discrepancy (MPP); +3° in one case, and +1° in the other two. As the AL increased further, the difference between the MPP and the AL decreased until, for one *S*, the MPP for an AL of 36° was also 36°; i.e., a zero discrepancy was preferred.

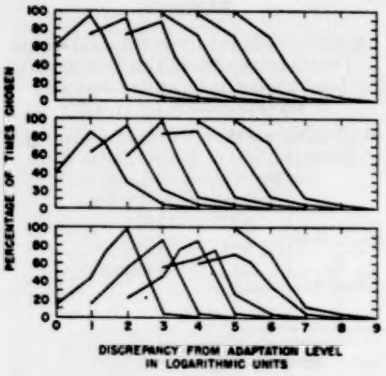

FIG. 3. Affect as a function of discrepancies of temperatures above the AL, when the AL is increased from 33° to 40°. Data taken from Table 3.

For the second *S*, the MPP was equal to a zero discrepancy at an AL of 38°. The third *S* did not reach an MPP of zero discrepancy until the AL had reached 40°. Above these points, the affect curve decreased in a linear fashion as discrepancy from AL increased further. With these three *S*s, it was impossible to establish complete adaptation above 38°. This was defined as not being able to report the neutral sensation in the AL bucket. Above this value, all *S*s reported a feeling of warmth.

DISCUSSION

These results supply ample support for the discrepancy hypothesis as to the sources of affect in the temperature sense. For an AL set at or near skin temperature, affect increased and then decreased as a regular function of discrepancies from the AL when the discrepancies were both above the AL and below it. AL's were also established above normal skin temperatures in three *S*s up to 38°. Geldard (**4**, p. 220) reports that a temperature of 38° takes 4–5 min. to adapt, while in this experiment, *S*s rarely took that much time to report adaptation. Since *S*s knew the session would last only as long as the time required to make the specified number of judgments, in some cases they may have reported complete adaptation before it had actually occurred, to speed up a somewhat boring task. If this happened, it might explain the leveling out of the "butterfly" effect. It is interesting to speculate as to whether the butterfly would have remained intact if adaptation had been complete at higher temperatures. Following Hebb's theory, this should be expected, because once a cell assembly is established, a slight discrepancy from it should be pleasurable. But if the assembly is not fully organized (i.e., adaptation is incomplete), then the discrepancy is measured from whatever assemblies are organized. This implies that if adaptation to 40° is not complete, 41° is not a discrepancy of +1° but really +3° or +4°, depending upon the maximal temperature to which *S* is adapted. Accepting this reasoning, affect is produced according to the discrepancy hypothesis whenever adaptation is complete, and that in the temperature sense, there is a considerable range over which adaptable AL's can be established.

This experiment involved the actual adaptation of the sense organs, not the central tendency of a series of sensations over time as in many of Helson's examples. The discrepancy hypothesis is applicable to this latter case, where the cell assembly would be the pattern of central nervous system activity resulting from repeated stimulation of a sensory nerve. Alpert (**1**) demonstrated the establishment of an adaptation level defined as the summation of previous auditory stimulation. He chose a pattern of very unfamiliar beats, making up a unique rhythm, as a single stimulus, and presented it many times. He argued that, at first, it would be so discrepant from previous AL's for rhythm that it should be unpleasant. As it was repeated, the previous AL's would be modified to include this new pattern. In this way, the discrepancy

should decrease and the stimulus would become more pleasant. Eventually, the AL should become identical with the new pattern and the affect would be indifferent. Alpert verified these predictions.

The discrepancy hypothesis is not the only possible explanation of affect. Any of the various forms of a tension reduction theory (e.g., **8**) might be extended to explain affect by assuming that central nervous excitation is minimal when the end-organs are adapted, and that any change from adaptation will produce an increase in stimulation. If increases in stimulation set up a drive state which *S* tries to reduce (which would appear to be a behavioral definition of negative affect), then a zero discrepancy from the AL should be preferred over all larger discrepancies, according to this theory. This prediction is clearly not borne out by these data, where small discrepancies from AL are decidedly preferred to zero discrepancies. This implies that increases in stimulation can be pleasurable, i.e., sought in themselves and not simply as a means to anxiety reduction, a fact that some tension-reductionists (e.g., **3**, p. 5; **9**) are just coming to recognize. Olds' data (**10**) on the reinforcing effects of *increases* in septal stimulation also support this view. These data seem more in line with a two-factor theory of motivation: that individuals may approach "pleasurable" objects in addition to avoiding "unpleasurable" ones. The discrepancy hypothesis is an attempt to give an independent definition to what is pleasant or unpleasant, independent of the resulting approach or avoidance behavior.

Summary

The *Ss* adapted both hands to water at or near skin temperature, and then placed each in water of different temperatures, behaviorally making an affective response as to the more pleasant temperature by withdrawing immediately the less comfortable hand. The results showed that small discrepancies from the adaptation level of the hands were more preferred than either larger discrepancies or zero discrepancies. Adaptation levels were then established above normal skin temperature in three *Ss*. The affective function shifted from a "butterfly" to a monotonically decreasing curve when the adaptation level was increased sufficiently. The discussion pointed out some of the implications of these data for motivational theories.

References

1. Alpert, R. The perceptual determinants of affect. Unpublished master's thesis, Wesleyan Univer., 1953.
2. Atkinson, J. W. Explorations using imaginative thought to assess the strength of human motives. In M. R. Jones (Ed.), *The Nebraska symposium on motivation.* Lincoln: Univer. of Nebraska Press, 1954.
3. Farber, I. E. Anxiety as a learned drive state. In M. R. Jones (Ed.), *The Nebraska symposium on motivation.* Lincoln: Univer. of Nebraska Press, 1954.
4. Geldard, F. A. *The human senses.* New York: Wiley, 1953.
5. Hebb, D. O. *The organization of behavior.* New York: Wiley, 1949.
6. Helson, H. Adaptation level as frame of reference for prediction of psychological data. *Amer. J. Psychol.*, 1947, **60**, 1–29.
7. McClelland, D. C., Atkinson, J. W., Clark, R. A., & Lowell, E. A. *The achievement motive.* New York: Appleton-Century-Crofts, 1953.
8. Miller, N. E., & Dollard, J. *Social learning and imitation.* New Haven: Yale Univer. Press, 1941.
9. Myers, A. K. & Miller, N. E. Failure to find a learned drive based on hunger; evidence for learning motivated by "exploration." *J. comp. physiol. Psychol.*, 1954, **47**, 428–436.
10. Olds, J. Physiological mechanisms of reward. In M. R. Jones (Ed.), *The Nebraska symposium on motivation.* Lincoln: Univer. of Nebraska Press, 1955.
11. Young, P. T. The role of hedonic processes in motivation. In M. R. Jones (Ed.), *The Nebraska symposium on motivation.* Lincoln: Univer. of Nebraska Press, 1955.

(Received January 2, 1958)

Journal of Experimental Psychology
Vol. 56, No. 4, 1958

ANAGRAM SOLUTION TIMES: A FUNCTION OF LETTER ORDER AND WORD FREQUENCY[1]

M. S. MAYZNER AND M. E. TRESSELT

Bell Telephone Laboratories and New York University

The present study is concerned with the effect of letter order and word frequency on anagram solution times. Anagram letter order refers to the rearrangement of the letters from their correct position in the word. Ten particular rearrangements or letter orders were selected for study. Five letter orders, designated "easy," were picked on a basis of maximizing the "similarity" of word letter order and anagram letter order. Thus, if word letter order is numbered 12345, an easy anagram letter order might be 23451, i.e., the first letter of the word is the last letter of the anagram, the second letter of the word is the first letter of the anagram, etc. Five other letter orders, designated "hard," were picked on a basis of minimizing the "similarity" of word letter order and anagram letter order, e.g., a hard anagram letter order might be 52413. Since with an easy letter order fewer movements of letters are required to place them in their proper position for word solution than with a hard letter order, it was hypothesized easy letter order anagrams would have faster solution times than hard letter order anagrams.

The second variable selected for study, word frequency as measured by Thorndike-Lorge frequency counts (6), has been found to have an appreciable effect on word recognition times (5). It is reasonable to expect, then, that it might have an analogous effect on anagram solution times. Therefore, it was hypothesized that anagrams whose solutions are words having high Thorndike-Lorge frequency counts would have faster solution times than anagrams whose solutions are words having low frequency counts.

[1] The authors wish to express their thanks to Paul S. Olmstead of Bell Telephone Laboratories for his helpful advice on the statistical portions of this paper.

METHOD

Subjects.—Forty *Ss* from introductory psychology classes at New York University were employed in the experiment, 20 in each of the two major conditions.

Material.—The stimulus material consisted of 20 five-letter anagrams typed in caps on plain white 3 × 5-in. index cards. The 20 words from which the anagrams were constructed were selected from the Thorndike-Lorge word list to provide four groups of five words each, differing in frequency of occurrence as follows: Group VF (very frequent) was characterized by words occurring 100 times or over per million and were *chair, sugar, train, party,* and *labor*; Group F (frequent) by words occurring at least 50 times per million and not so many as 100 times per million and were *beach, model, clerk, fault,* and *giant*; Group I (infrequent) by words occurring at least once per million and not as many as twice per million and were *patio, cobra, roach, baton,* and *jaunt*; and Group VI (very infrequent) by words occurring less than once per million but more than once per four million and were *tango, groin, peony, ghoul,* and *triad.*

Experimental design.—The *Ss* were assigned randomly to the two major conditions which involved presenting the same 20 words in either the easy or the hard letter orders. The five easy letter orders were: 12354, 23451, 51234, 34512, 45123. The five hard letter orders were: 14253, 25314, 52413, 31425, 42513. In brief, 20 *Ss* received 20 words employing five different easy letter orders each repeated four times for the four frequency groups, while another 20 *Ss* received the same 20 words employing five different hard letter orders each repeated four times for the four frequency groups.

Procedure.—The *Ss* were told that a five-letter anagram is simply five scrambled letters

which when properly rearranged will make a single five-letter word. After being shown an example, they were informed that they would have 20 anagrams to solve, and if they could not find a solution in 4 min., the correct solution would be given to them before going on to the next one. Solutions were given by *S* verbally (i.e., *S* had no paper or pencil) and solution times were recorded with a stop watch. All *S*s were tested individually.

Results

Because of the nature of the distributions of the five time scores for each word group (VF, F, I, and VI), letter order (Easy and Hard) combination for each *S*, the *S*'s median time score for each combination was selected as most representative of his performance for analysis.

Table 1 gives the major results of the study showing means, *SD*'s, medians, and ranges for the time scores. Both means and medians seveal a steady increase in anagram rolution times going from the Easy Order-VF condition to the Hard Order-VI condition, except for one slight reversal between the VF and F word groups under the easy letter order condition. Within this design several comparisons are possible. To evaluate the effect of the letter order variable, comparisons of the easy versus hard letter order conditions for each of the four word frequency groups, were carried out employing the Mann-Whitney *U* test (**4**). All comparisons yielded probability values well below the .001 level.

In evaluating the effect of the word frequency variable, two separate analyses were made. Under each letter order condition, the Friedman Two Way Analysis of Variance by Ranks Test (**3**) was employed. The probability value associated with the word frequency groups comparison under the easy letter order condition was less than the .01 level, while the equivalent probability value yielded for the word frequency groups comparison under the hard letter order condition was less than the .001 level.

TABLE 1

Anagram Solution Times (in Sec.) for Letter Order and Word Frequency Conditions (Based on Medians)

Group	Mean	*SD*	Median	Range
Easy Order				
VF	9.7	5.7	8.5	2–19
F	7.3	4.2	6.0	2–15
I	20.5	23.8	12.0	4–89
VI	46.8	66.3	13.0	2–237
Hard Order				
VF	51.0	58.6	25.0	5–240
F	55.7	67.4	32.5	7–240
I	116.5	83.0	96.0	18–240
VI	135.0	91.6	151.0	21–240

Note.—Time scores on which means, *SD*'s, medians, and ranges are calculated are the median solution times for each frequency group of five anagrams.

Discussion

The results of this study demonstrate the highly significant effect of letter order and word frequency variables on anagram solution times. A suggested theoretical framework designed to account for the effect of these variables would involve the following considerations. An anagram constitutes a problem whose stimulus properties tend to evoke a variety of implicit responses on the part of *S*. When the number of letter rearrangements or letter movements is few to go from anagram to word solution, as in the easy letter order condition, fewer implicit responses or hypotheses presumably are needed to solve the anagram than when many letter movements are needed. It becomes reasonable to expect, then, that the greater the number of implicit responses that need to be made, the longer the anagram solution times will be.

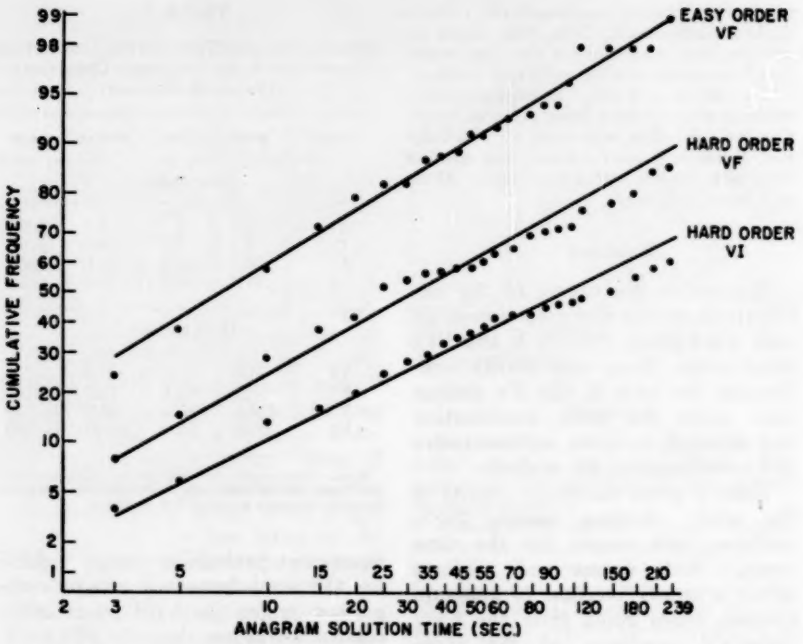

FIG. 1. A plot showing the relationship between anagram solution time in seconds (log scale) and the cumulative frequency distribution (probability scale) for a sample of 100 anagrams for the Easy Order-VF, Hard Order-VF, and Hard Order-VI conditions.

Further, in the mechanism governing the generation of *S*'s implicit responses, if the solution to an anagram is a word having a high frequency of occurrence in the language, it will probably be relatively high in *S*'s response repertoire, and therefore possess a greater potential for evocation as an early implicit response than a word having a low frequency of occurrence. Thus, faster anagram solution times should be associated with words having high frequency of occurrences (i.e., greater evocation potential in *S*'s response repertoire) as opposed to words having low frequency of occurrences.

If it is granted that anagram solving processes involve, in part at least, the production of a series of implicit responses prior to the final correct solution, then the question arises, "How is this production process distributed through time?" No direct answer is available from the present study; however, an indirect attempt to examine this production process was carried out as follows. Each of the eight word group, letter order combinations yield, if all anagram solution times are employed, a sample of 100 solution times. A cumulative plot was made for each group or sample of 100 anagram solution times on logarithmic probability paper, and the results for three of the groups are shown in Fig. 1. The remaining five groups yield essentially similar straight line distributions paralleling and interspersed with those shown. The lines drawn are fitted visually.

A possible inference that can be drawn, then, from the results shown in Fig. 1, and which is in accord with the general

mathematical properties of such logarithmic plots (**2**), is that, the more time *S* takes up to any given point without solution, the more time he will probably take beyond that point before a successful solution occurs. Basically, it would appear that some given number of implicit responses or hypotheses are produced rather quickly, and if one of these matches the actual word, the anagram is solved, with the probability of a correct match being primarily a function of the difficulty level of the anagram. Thus, for example, within 30 sec. 82% of the anagrams are solved in the Easy Order-VF condition, while only 28% are solved in the Hard Order-VI condition. If on the other hand, the anagram is not solved quickly we know proportionally more time will be taken before solution, and this would seem to imply either, (*a*) a less rapid production of new implicit responses or hypotheses through time, or (*b*) possibly extensive retesting of old hypotheses through time. Work by Bousfield and Sedgewick (**1**) is highly relevant to this issue. They asked *S*s to name as many cities in the United States as they could, or to make as many words as possible drawing from a given pool of letters (similar to anagram solving, but different in that the words could vary in length and many different words could be constructed from the letter pool given). The results clearly revealed that the production of responses was rapid at first, progressively slowing down through time. The results of the present study, as analyzed in Fig. 1, would seem to imply that the production of implicit responses or hypotheses to an anagram problem situation can be described by the same law as that governing the production of explicit responses in verbal recall and word problem situations.

Summary

A study was made of the effect of letter order and word frequency variables on anagram solution times. Two levels of anagram letter order and four levels of word frequency were examined. Letter order was varied by maximizing (easy order) or minimizing (hard order) the "similarity" of anagram letter order to word letter order. Word frequency was varied by selecting words varying in their Thorndike-Lorge frequency counts from very frequent to very infrequent.

The results clearly showed the highly significant effect of these two variables on anagram solution times. Specifically, hard letter orders and low word frequency counts both produced marked increases in anagram solution times. Further analysis revealed the logarithmic nature of the distribution of anagram solution times, indicating that the more time *S* takes up to any given point without solution, the more time he will probably take beyond that point before a successful solution occurs.

References

1. Bousfield, W. A., & Sedgewick, C. H. W. An analysis of sequences of restricted associative responses. *J. gen. Psychol.*, 1944, **30**, 149–165.
2. Cramer, H. *Mathematical methods of statistics.* Princeton: Princeton Univer. Press, 1946.
3. Friedman, M. The use of ranks to avoid the assumption of normality implicit in the analysis of variance. *J. Amer. statis. Ass.*, 1937, **32**, 675–701.
4. Mann, H. B., & Whitney, D. R. On a test of whether one of two random variables is stochastically larger than the other. *Ann. math. Statist.*, 1947, **18**, 50–60.
5. Solomon, R. L., & Howes, D. H. Word frequency, personal values, and visual duration thresholds. *Psychol. Rev.*, 1951, **58**, 256–270.
6. Thorndike, E. L., & Lorge, L. *The teacher's word book of 30,000 words.* New York: Bureau of Publ., Columbia Univer., 1944.

(Received for early publication
May 7, 1958)

Journal of Experimental Psychology
Vol. 56, No. 4, 1958

ABSOLUTE JUDGMENTS OF STATIC PERSPECTIVE TRANSFORMATIONS

RAYMOND C. SIDORSKY

Electric Boat Division, General Dynamics Corporation

In a recent study Gibson and Gibson (6) demonstrated that a continuous perspective transformation in the optic array can lead to the perception of rigid motion. The continuous compression of a pattern in one meridian was seen as a changing slant of a rigid surface. The impression of changing slant was obtained with regular (quadrilateral) vs. irregular (ameboid) and textured vs. nontextured stimuli. Essentially similar linear functions relating the estimated slant to the slant in the optic stimulus were obtained in all four cases.

A control group was shown a single motionless example of each of the four patterns slanted 60° away from the frontal plane. In this case, only the regularly patterned stimuli appeared slanted. The authors attributed this result to the fact that the regular pattern had a unique member—the square—with immediately identifiable members. The irregular pattern lacked such a member. Since judgments of the extent of the slant with moving stimuli were equally accurate for both regular and irregular patterns, it is apparent that motion in the stimulus is necessary to induce any perception of depth with irregular and/or unfamiliar patterns. Motion in the stimulus apparently also has a significant effect upon *O*'s judgment of the extent of slant of regular patterns. Although *O*s saw the motionless pattern as slanted away from frontal plane, the mean estimates of the extent of the slant were quite inaccurate. For example, with the regular patterns a 60° slant in the optic stimulus was judged, on the average, as 61° (*SD* about 6°) with moving stimuli, but only 24° (*SD* about 12°) with motionless stimuli. Thus, it would appear that *accurate* estimation of the extent of the apparent slant of a transformed perspective image depends, at least in part, upon motion in the optic stimulus.

However, the authors were primarily interested in the elicitation rather than the accuracy of depth perception; consequently, the experimental design was such that the specific contribution of motion in the stimulus to accurate depth perception could not be determined. For one thing, *O*s in the control group were naive with regard to the object transformed and the transformation sequence, i.e., they were never shown the original pattern (square) in the frontal plane or other members of the transformation family. Furthermore, they had no opportunity to practice. The aim of the present study was to determine the degree to which accurate estimates of the extent of a perspective transformation depend upon motion in the stimulus per se. Perspective images of a regular (quadrilateral) pattern were used. The *O*s were told that the display represented a perspective image of a rigid surface. They were also shown representative members of the family of static perspective transformations that constituted the stimulus series. Subsequently, the accuracy with which such sophisticated *O*s could make absolute judgments of the

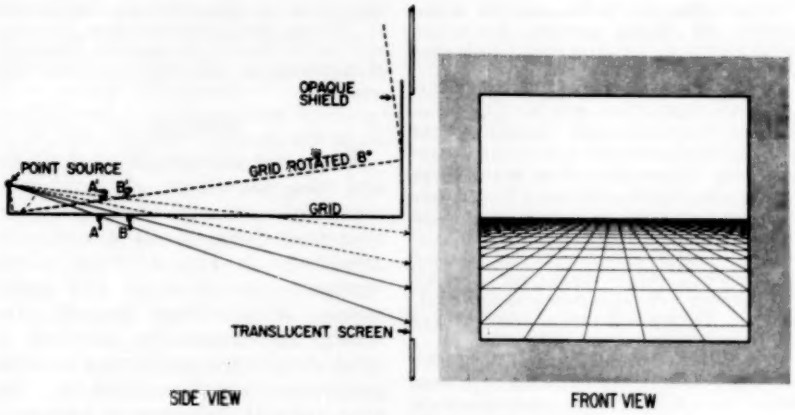

FIG. 1. Schematic view of shadow transformer.

extent of the apparent rotation of a motionless perspective image was measured. At no time did *O*s see the optic stimulus in motion.

The problem is of interest in another context. Gibson's formulations (**3, 4, 5, 6, 7,**) have provided the theoretical basis for research and development of a new display system—generally designated the "contact analog."[1] The contact analog is essentially a two-dimensional projection of a model surface which undergoes continuous perspective transformations corresponding to all six degrees of freedom associated with movement in three-dimensional space. Gibson's hypothesis that optic stimuli need only correlate with, not necessarily replicate, direct experience has led to the further hypothesis that more effective control of the orientation and movement of objects or vehicles in space can be effected through the use of a display which provides an analog (correlate) of contact (direct experience) flight.

[1] Research and development of the contact analog has been sponsored by the Office of Naval Research.

METHOD

The generation of a perspective transformation can be described in terms of a moving surface and a fixed *O*, or alternately, of a fixed surface and a moving *O*. In Gibson's work *O* has generally been considered fixed and the surface in motion. In the contact analog situation the surface is considered fixed and *O* in motion. For a given transformation the optic stimulus is the same in either case. The present study was carried out in the latter context; *O*s were told that, although they would actually remain level, they were to assume that the changes in the configuration of the display resulted from their own motion. The transformation sequence used was pitch, i.e., change in the optic stimulus corresponding to change in the configuration of a patterned surface seen by an *O* undergoing apparent rotation in the medial plane.

Apparatus.—The shadow transformer used to generate the display was similar in principle to that used by Gibson and Gibson (**6**). It is shown schematically in Fig. 1. The ray sheaf of a 25-w., .5-mm. filament point source was passed through a 30-in. sheet of clear plastic upon which an opaque grid pattern of 1/4-in. outline squares had been inscribed. The resultant shadow pattern fell on a vertical translucent screen seen through an 18 × 20-in. aperture in a large sheet of plywood. The screen was located 30 in. from the point source. The *O* was seated on the opposite side of the display screen. With the grid in a horizontal position, the shadow pattern seen by *O* formed a perspective image of a flat surface covered with squares and extending to an artificial

"skyline" (Fig. 1). At this point the skyline bisected the display aperture. An opaque shield fixed to the end of the grid holder blocked all light above the skyline.

The grid was rotated tangent to a circle with a 1-in. radius centered at the point source filament. Counterclockwise rotation of the grid produced perspective transformations corresponding to those which would be seen by an *O* undergoing clockwise rotation. The patterns displayed were correlates of those which would be seen by an *O* seated 30 in. from an 18 × 20-in. window, rotated (along with the window) about the nodal point of the eye, and suspended 200 ft. above a flat surface covered with 50-ft. squares. The apparent distance between *O* and the skyline was approximately 2.3 miles.

The *E* varied the angle of the grid manually; referring to a fixed scale indicating pitch angle from 0° to −40° in 2° steps. A friction brake held the grid firmly in position during the stimulus exposure. Stimuli were presented by removing a solenoid-operated shutter in front of the point source. Two Cramer timers wired for automatic recycling controlled the duration of the stimulus exposure and stimulus change periods.

Observers.—Twenty-five naval officers were used as *O*s. Since all had had some experience flying in small aircraft, no difficulty was encountered in explaining the nature of the perspective transformation displayed.

Procedure.—The *O*s were seated 30 in. from the display screen in a dimly illuminated room. Their seated height was adjusted so that with the 0° image the skyline was at eye level. The *E* emphasized that, although they actually would remain level, *O*s were to assume that the differences in the configuration of the display resulted from their own apparent rotation about the horizontal axis while viewing a fixed surface. The *O*s were then shown static end products of transformations corresponding to pitch changes from 0° to −10° in 2° steps, from −14° to −40° in 4° steps, and returning, from −40° to 0° in 10° steps. Each stimulus appeared for 4 sec. followed by a 4-sec. interval during which the display was changed. The same time intervals were used during the remainder of the session. During the interval between trials the display screen was illuminated at a brightness level approximately equal to that of the stimulus. This was done to eliminate disturbing brightness contrasts after each stimulus presentation.

The method of absolute judgments (2) was used to collect and analyze data. The *O*s were instructed to judge and verbally designate their apparent "pitch angle" to the nearest 2° step. Each "pitch angle" was displayed 10 times to each *O* during the session. A different unsystematic order of presentation was used for each *O*. A 5-min. rest interval was given after one-third and two-thirds of the trials were completed. No knowledge of results was given during the session.

Results

The mean estimate of pitch angle and the variability of judgments associated with each of the perspective transformations are shown graphically in Fig. 2. Each point represents the mean of 250 judgments. It is evident that *O*s can readily and accurately interpret a series of static end products of suitable perspective transformations in the optic array as differences in apparent pitch. Constant errors ranged from an overestimation of 4.22° at −16° to an underestimation of 2.10° at −40° with an over-all mean of 1.91°. Variable errors based on pooled data ranged from an *SD* of 2.01° at 0° to an *SD* of 5.32° at −34° with an over-all mean *SD* of 4.11°.

The mean constant and variable errors listed above are appropriate as indices of over-all performance with a hypothetical contact analog type display. However, further analysis of the scores associated with the

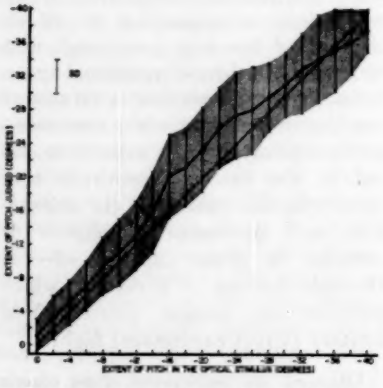

Fig 2. Mean estimate and variability of judgments associated with each pitch angle.

individual pitch angles is required since not all of these scores were based upon judgments with respect to a single perceptual cue. In a sequence of transformations starting at 0°, the distance between the skyline—present in a portion of the stimulus series—and the edge of the display aperture serves as an additional cue to the extent of the transformation. The effect of this additional cue upon *O*s' performance is apparent in Fig. 2. In the portion of the stimulus series in which the skyline was present (0° to −12°), pitch angles were *consistently* overestimated at a level slightly in excess of one degree (mean = 1.17°). Beginning at −14°, however, a different function was obtained. The constant error increased abruptly and changed from an overestimation of 4.22° initially to an underestimation of 2.10° at the end of the series. The mean constant error was 2.25°. Variability also showed a similar but less well defined change in the neighborhood of −14°. The estimates obtained in the −14° to −40° portion of the stimulus series thus represent more closely the accuracy with which *O*s can judge the extent of the apparent rotation of a perspective image based solely upon differences in the configuration of a motionless pattern.

The variability between *O*s and the average of the *SD*'s within *O*s associated with each transformation are shown in Fig. 3. In both cases variability increased as a function of the distance from the ends of the stimulus series.

An information analysis of *O*s' responses was performed using the procedure described by Garner and Hake (2). The average information transmitted by each *O* was computed separately. Information transmitted ranged from 1.92 bits to 3.32 bits per *O*, averaging 2.52 bits for the group

The antilog$_2$ of the average information in bits transmitted in an absolute judgment represents the approximate number of stimulus categories that *O*s can discriminate without error (1, 2). On this basis, the results of the present study indicate that, within the range from 0° to −40°, *O*s can differentiate approximately six pitch angles based on observation of the static end products of a perspective transformation.

As indicated earlier, Gibson and Gibson (6) found that estimates of the slant of a static perspective image were relatively inaccurate compared with those obtained with a continuous transformation. The results of the present study indicate that the gross underestimation of the slant of static perspective transformations of a regular pattern found by Gibson and Gibson (6) is not a consequence of the lack of motion in the stimulus. If, as in the present case, *O*s are informed that the differences in the configuration of a display are the result of a perspective transformation of a rigid surface, a series of line patterns never seen in motion can nonetheless carry a transformation sequence.

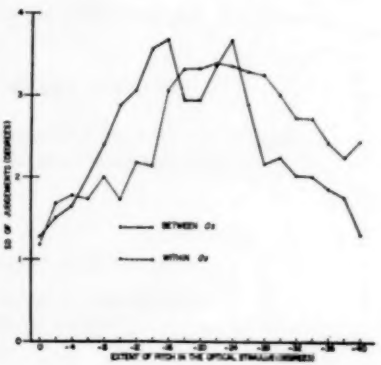

FIG. 3. Variability within and between *O*s associated with each pitch angle.

Summary

Static perspective images corresponding to a flat grid-patterned surface viewed at various pitch angles were projected onto a vertical screen. The 21 perspective patterns which would result from the rotation of the *O* in 2° steps from 0° to −40° were displayed. Representative members of the family of static perspective transformations that constituted the stimulus series were shown to 25 naval officers who were then asked to judge and verbally designate the extent of their apparent rotation in the medial plane. The method of absolute judgments was employed. Each *O* made 10 judgments of each of the 21 stimuli.

The *O* readily and accurately interpreted differences in the extent of a perspective transformation as differences in their own pitch angle with respect to a fixed surface. Constant errors ranged from −2.10° to 4.22° with an over-all mean of 1.9°. Stimuli which lacked a visible skyline were judged less accurately (mean CE 2.25°) than those with a skyline. Information transmitted averaged 2.52 bits for the group indicating a capability of differentiating without error approximately six pitch angles between 0° and −40°.

References

1. Alluisi, E. A., & Sidorsky, R. C. The empirical validity of equal discriminability scaling. *J. exp. Psychol.*, 1958, **55,** 86–95.
2. Garner, W. R., & Hake, H. W. The amount of information in absolute judgments. *Psychol. Rev.*, 1951, **58,** 446–459.
3. Gibson, J. J. The perception of visual surfaces. *Amer. J. Psychol.*, 1950, **63,** 376–384.
4. Gibson, J. J. *The perception of the visual world.* Boston: Houghton Mifflin, 1950.
5. Gibson, J. J. The visual perception of objective motion and subjective movement. *Psychol. Rev.*, 1954, **61,** 304–314.
6. Gibson, J. J., & Gibson, E. J. Continuous perspective transformations and the perception of rigid motion. *J. exp. Psychol.*, 1957, **54,** 129–138.
7. Gibson, J. J., Olum, P., & Rosenblatt, F. Parallax and perspective during aircraft landings. *Amer. J. Psychol.*, 1955, **68,** 372–385.

(Received for early publication
May 16, 1958)

ZEITSCHRIFT FÜR
EXPERIMENTELLE UND ANGEWANDTE PSYCHOLOGIE

Organ der Deutschen Gesellschaft für Psychologie

Herausgeber für Deutschland

PROF. DR. PH. LERSCH, München; PROF. DR. F. SANDER, Bonn

Herausgeber für Österreich	Herausgeber für die Schweiz	Redaktion
PROF. DR. H. ROHRACHER, Wien	PROF. DR. R. MEILI, Bern	DR. C. J. HOGREFE, Göttingen

Band V/1

GOTTFRIED EWALD, **Psychologie, Psychotherapie und Psychiatrie**

ROLF AHRENS, **Zur Bedeutung der Psychologie für das Aphasieproblem**

ERICH MITTENECKER, **Die Analyse "zufälliger" Reaktionsfolgen**

ANNELIESE KRANTZ-GROSS und HANS-JOACHIM HAASE, **Psychomotorik des Griffdrucks bei endogenen Depressionen**

GUSTAV A. LIENERT, **Ein Formlegetest als Prüfmittel der praktischen Intelligenz**

EBERHARD ULICH, **Neue Erfahrungen mit dem Pauli-Test**

ANNE-MARIE TAUSCH, **Empirische Untersuchungen über das Verhalten von Lehrern gegenüber Kindern in erziehungsschwierigen Situationen**

All articles are provided with summaries in English and French.

The journal appears quarterly and contains 160 pages per number.
Subscription is DM 64.—per volume.

VERLAG FÜR PSYCHOLOGIE • DR. C. J. HOGREFE • GÖTTINGEN
(Brentanoweg 10, Germany)

PSYCHOLOGICAL MONOGRAPHS

The monographs listed below are still in print—but for how long?*

ADELBERT AMES, JR. *Visual Perception and the Rotating Trapezoidal Window.* No. 324. *$1.00*

JANE SHOHL. *Effects of Exposure to Sound on Discrimination Performance in the Rat.* No. 329. *$1.00*

PHILIP WORCHEL. *Space Perception and Orientation in the Blind.* No. 332. *$1.00*

P. R. WENDT. *Development of an Eye Camera for Use with Motion Pictures.* No. 339. *$1.00*

MONCRIEFF SMITH and EDNA A. WILSON. *A Model of the Auditory Threshold and Its Application to the Problem of the Multiple Observer.* No. 359. *$1.50*

M. R. ROSENZWEIG and W. A. ROSENBLITH. *Responses to Auditory Stimuli at the Cochlea and at the Auditory Cortex.* No. 363. *$1.00*

M. B. JONES. *An Experimental Study of Extinction.* No. 369. *$1.00*

G. F. RABIDEAU. *Differences in Visual Acuity. Measurements Obtained with Different Types of Targets.* No. 395. *$1.00*

---

* A selection from Volumes 65–69. (Write for complete list of in-print items.)

Order *now* from: **American Psychological Association**
1333 Sixteenth Street, N. W.
Washington 6, D. C.